Identification of Common and Distinct Epigenetic Reprogramming Properties of Core-binding Factor Fusion Proteins

(Justin) Ching Ting Loke

A Thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences,

College Of Medical and Dental Sciences,

University of Birmingham

November 2016

Abstract

RUNX1, also known as CBFα, is a master regulator of haematopoiesis. In Acute Myeloid Leukaemia (AML) it is frequently disrupted by translocations to different epigenetic regulators, resulting in the expression of core-binding factor fusion proteins.

We compared the chromatin landscape of t(8;21) and t(3;21) AML which express RUNX1-ETO and RUNX1-EVI1, respectively. We found that the diverse clinical outcomes of patients with these two forms of AML are reflected in fundamental differences in gene expression and chromatin landscape. Despite both fusion proteins sharing a RUNT DNA binding domain, we show that RUNX1-EVI-1 targets a more immature stem cell-related gene expression program of genes as compared to RUNX1-ETO.

Despite the differences in the epigenomic landscape of t(3;21) and t(8;21) leukaemia, knockdown of either core-binding factor fusion protein activates a common myeloid differentiation program involving up regulation of C/EBPα. By blocking C/EBPα DNA binding through a dominant negative partner, we showed that this factor is required for the downstream effects of RUNX1-EVI-1 knockdown.

Even in the continued presence of RUNX1-EVI-1, ectopic expression of C/EBPα is sufficient to initiate myeloid differentiation in t(3;21) cells. Overall, this suggests that deregulation of C/EBPα is a common pathway in the development of both t(8;21) and t(3;21) AML.

This work is dedicated to my wife, Rebecca, and our daughter, Bessie.

Acknowledgments

I would like to thank my supervisor Prof Constanze Bonifer for this opportunity to work with her team. She has been a constant source of wisdom and infective enthusiasm.

Dr Salam Assi performed the bioinformatics analysis which has brought this story alive; I would like to thank her for her perseverance with this project, and her insightfulness. I am also indebted to Dr Anetta Ptasinska and Dr Maria "Rosie" Imperato for teaching me many of the techniques used in this thesis. Furthermore, Dr Ptasinska generated the majority of the t(8;21) AML data used in this thesis. Dr Imperato was an important colleague in much of the lentiviral work and DNase I hypersensitivity site mapping experiments. I would also like to thank the rest of the Bonifer-Cockerill lab for their support.

Prof Peter Cockerill and Dr Manoj Raghavan provided helpful advice. Prof Olaf Heidenreich designed the siRNA against RUNX1-EVI-1, on which much of this work relies, and he has also been a source of fruitful discussions. I would like to thank his lab, and especially Dr Natalia Martinez-Soria, for helping set up the RUNX1-EVI-1 knockdown. I would also like to thank Prof Ruud Delwel for advice on the RUNX1-EVI-1 ChIP-seq and for patient samples.

I would like to thank 'NHS Blood and Transplant' for reagents, and colleagues at the Department of Haematology, Queen Elizabeth Hospital, Birmingham for their support. I would also like to thank both the high-throughput sequencing facility at Biosciences, University of Birmingham and also the facility run by Dr Andrew Beggs, for their efficiency.

This work was funded through a Kay Kendall Leukaemia Fund junior clinical fellowship, of which I am grateful.

Finally, I am indebted to my family for their support during these studies, and especially to my parents for guiding me to this stage.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

5mC- 5-methyl-cytosine

AGM-Aorta-Gonadal-Mesonephros

AML- Acute Myeloid Leukaemia

ASXL1- Additional Sex Combs Like 1, Transcriptional Regulator

ATAC-seq-assay for transposase accessible chromatin using sequencing

ATP - Adenosine Triphosphate

BMI1 - B-lymphoma insertion region 1

BSA-Bovine Serum Albumin

CBF-Core-Binding Factor

CBP - CAAT-Binding Protein

CD - Cluster of Differentiation

cDNA - complementary DNA

CEBP- CCAAT-Enhancer-Binding Protein

ChIP-seq-Chromatin Immunoprecipitation with high-throughput sequencing

CLP- Common Lymphoid Progenitors

CML-Chronic Myeloid Leukaemia

CMP- Common Myeloid Progenitors

CO_2 - Carbon Dioxide

CO-IP- Co-immunoprecipitation

CTD - C-terminal Domain

DHS- DNase Hypersensitive Site

DMEM- Dulbecco Modified Eagle Media

DMSO – Dimethyl Sulfoxide

DNA – Deoxyribonucleic acid

DNCEBP- Dominant negative CEBP

DSG- Di(N-succinimidyl) glutarate

EDTA- Ethylene diamine tetra acetic acid

eGFP- enhanced green fluorescent protein

EGTA – Ethylene glycol tetra acetic Acid

ETO - Eight Twenty-one Oncogene

EZH2 - Enhancer of Zeste

FACS- Fluorescence Activated Cell Sorting

FCS- Fetal Calf Serum

FLT3 - FMS Like Tyrosine Kinase

G-CSF - Granulocyte Colony Stimulating Factor

GFP - Green Fluorescent Protein

GMP- Granulocyte Monocyte Progenitors

GSEA- Gene Set Enrichment Analysis

HAT- Histone Acetylase

HDAC- Histone Deacetylase

HSC- Haematopoietic Stem Cell

IDH- Isocitrate dehydrogenase

IL – Interleukin

IMDM- Iscove Modified Dulbecco Media

JNK –Jun Kinase

KCl – Potassium Chloride

KD –Knock-Down

LiCl - Lithium Chloride

LMPP- Lymphoid Primed Multipotent Progenitors

LSC- Leukaemic Stem Cell

MACS- Magnetic activated cell sorting

MDS- Myelodysplastic Syndrome

MEP- Megakaryocyte Erythrocyte Progenitors

MgCl2 - Magnesium Chloride

MOI- Multiplicity of Infection

MPP- Multipotent Progenitors

NaCl - Sodium Chloride

NGS- Next Generation Sequencing

NHR- Nervy Homology Region

NHSBT- NHS Blood and Transplant service

NK- Natural Killer Cells

PBS- Phosphate buffered saline

PBSC-Peripheral Blood Stem Cell

PCA-Principle Component Analysis

PI - Propidium Iodide

PIC- Protease Inhibitor cocktail

RPMI-Roswell Park Memorial Institute

RNAPII- RNA polymerase II

RNA - Ribonucleic acid

RNA-seq- RNA sequencing

RT-qPCR- Quantitative reverse transcription PCR

SDS - Sodium Dodecyl Sulphate

siRNA - Small inhibitory Ribonucleic Acid

STAT – Signal Transducer and Activator of Transcription

SWI/SNF- Switch/Sucrose Nonfermentable

TAE – Tis-Acetate-EDTA

TBP –TATA Binding Protein

WHO – World Health Organisation

Chapter 1. Introduction

1.1 Haematopoiesis

Haematopoiesis is the process by which cells of the blood system develop. Haematopoietic stem cells (HSCs) are largely quiescent, dividing infrequently, but have self-renewal mechanisms that allow them to persist over the lifetime of an organism to replenish more rapidly dividing pools of downstream progenitors (Wilson et al., 2008, Sun et al., 2014, Busch et al., 2015). This process is carefully regulated by homeostatic mechanisms to allow the organism to respond to stresses such as infection and bleeding but rarely result in uncontrolled proliferation (Essers et al., 2009, Baldridge et al., 2010). HSCs are also multi-potent which allows differentiation into all the lineages of the haematopoietic system (Spangrude et al., 1988). This section will detail how the haematopoietic system emerges and is subsequently regulated.

1.1.1 The emergence of haematopoietic stem cells

HSCs emerge from the haemogenic endothelium that lines the surface of the major arteries, principally, in the aorto-gonadal mesonephros (AGM), before maturing in the fetal liver and finally, implanting in the bone marrow. These definitive HSCs are capable of maintaining an adult haematopoietic system. In contrast, prior to this time, cells from the yolk sac provide early haematopoietic precursors to sustain the developing embryo, but are unable to survive in the adult haematopoietic system (Cumano et al., 1996, Cumano et al., 2001). The AGM was shown to be the site where definitive HSCs emerge by transplanting tissue from this region into lethally irradiated mice where long-term haematopoiesis was successfully reconstituted (Medvinsky and Dzierzak, 1996, Medvinsky et al., 1993, Muller et al., 1994). Later, the sites of HSC emergence in situ were identified through gene-labelling experiments (Zovein et al., 2008) and visualised through confocal imaging of dissected murine embryos (Boisset et al., 2010). After this emergence at day 10.5 in mice, HSCs migrate through the circulatory system to develop further in the foetal liver, to finally reside in the

bone marrow (Gekas et al., 2005, Christensen et al., 2004, Morrison et al., 1995).

1.1.2 Hierarchical relationship of blood cell development

HSC differentiate into intermediate progenitors which display increasingly restricted lineage potential (figure1-1). This developmental process has been delineated through the use of labelling cell surface markers by antibody staining. Bone marrow subpopulations were isolated by magnetic beads or fluorescent activated cell sorting (FACS), and their function subsequently identified by transplantation into lethally irradiated mice. Initial work identified a population of HSCs (Spangrude et al., 1988, Muller-Sieburg et al., 1986) capable of differentiating into all lineages. Common lymphocyte progenitors (CLP) which only produce lymphocytes (Kondo et al., 1997) and the existence of common myeloid progenitors (CMP) capable of producing granulocyte-monocyte progenitors (GMP) and megakaryocyte-erythrocyte progenitors (MEP) were later identified (Akashi et al., 2000).

The precise fate of CMPs and the origins of GMPs have been challenged. For example, some groups have shown that MEPs come direct from short-term HSCs, and not from CMPs, suggesting the existence of a separate lymphoid primed multi-potential progenitor (LMPP). LMPPs, expressing FLT3, were shown to be precursors to both the lymphoid (CLP) and myeloid (GMP) populations (Adolfsson et al., 2005). This version of the haematopoietic hierarchy was subsequently debated by further studies (Pronk et al., 2007, Arinobu et al., 2007) using other cell surface markers. The development of indexed FACS coupled with single-cell RNA-seq technology will likely result in improved understanding of the functional heterogeneity that resides in the early progenitor compartments (Paul et al., 2015). In the following parts of this section we will discuss the ways in which these developmental trajectories can be regulated.

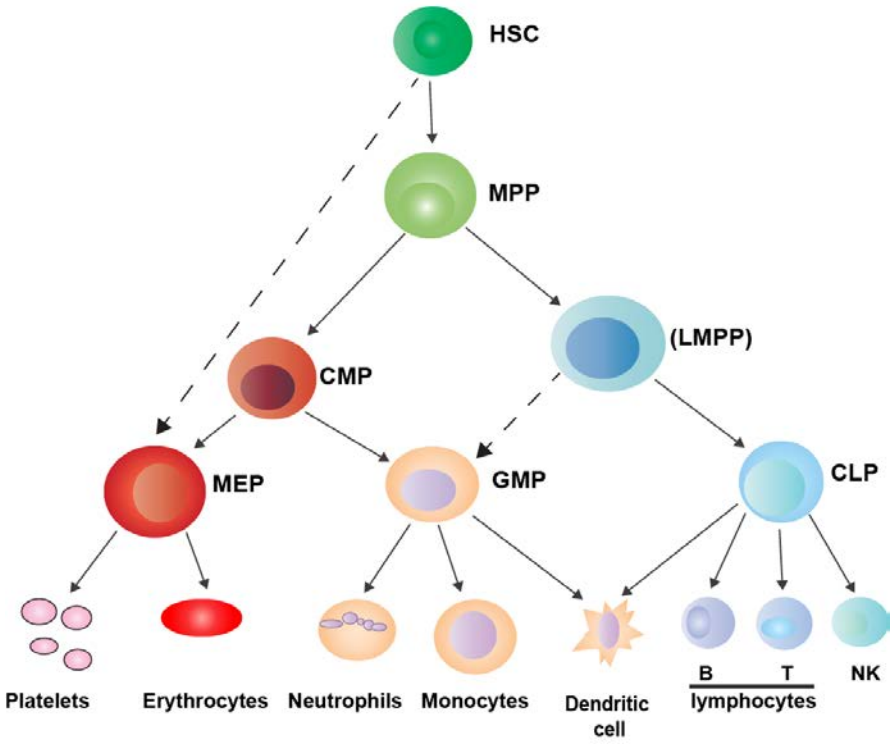


Figure 1-1 The haematopoietic hierarchy.
The development of haematopoietic cells from HSC to terminally differentiated effector cells. Dashed lines represent alternative differentiation pathway in the model proposed by Adolffsson et al., 2005, based on the presence of LMPPs.

1.1.3 Regulation of differentiation in haematopoietic cells

Regulation of haematopoiesis by signalling from microenvironment

A paradigm suggests that cytokines have a "permissive" role in haematopoiesis. In this paradigm, cytokine signalling only drives cell proliferation and survival, implying that differentiation to different cell fates is determined intrinsically by the expression of lineage specific transcription factors. A classical experiment in support of this view was conducted by Lagasse and Weissman in 1997. Mice were engineered that lacked the cytokine M-CSF and had low numbers of monocytes. Monocyte and neutrophil counts could be restored by permitting their survival through the transgenic expression of human BCL2, through the

endogenous *mrp8* promoter, a gene which is expressed from early myeloid progenitors onwards (Lagasse and Weissman 1997).

Another paradigm suggests that cytokines can be instructive in determining cell fate (Rieger et al., 2009). For example, the receptors for GM-CSF and M-CSF are expressed at low levels on primitive HSC but are down regulated upon differentiation to CLP. Human IL-2Rβ (hIL-2Rβ) can couple with mouse γc chain to form a receptor that exclusively responds to human IL-2 cytokine. A transgenic mouse model which expresses hIL-2Rβ constitutively was developed. Normal CLP do not express the IL-2 receptor. Unlike normal CLP, CLP expressing hIL2Rβ trans-differentiate into granulocytes and monocytes when exposed to hIL-2, through upregulation of GM-CSF and M-CSF receptors (Kondo, Scherer et al. 2000).

However, in contrast to knock out of transcription factors, knock out of components of signalling transduction pathway often results in an incompletely penetrant phenotype, arguing against the cytokine-instructive paradigm. For example, although death and severe anaemia results in erythropoietin receptor deficient mice, early erythroid progenitors still exist in their bone marrow and erythropoiesis can be detected in the yolk sac (Kieran, Perkins et al. 1996). In part this may be due to the compensatory effects of thrombopoietin (TPO) through the receptor c-MPL, as in vitro, burst forming unit erythroid (BFU-E) colonies can be formed when liver and yolk sac extracted cells were cultured with recombinant TPO.

Taken together, this suggests a pleotropic role for cytokine signalling in haematopoiesis with the ability for cytokines to deliver an instructive signal for differentiation in specific developmental contexts.

Regulation of haematopoiesis by transcription factors

The importance of transcription factors in the regulation of haematopoiesis was established early on, through the identification of transcription factors which are essential for the normal development of the haematopoietic hierarchy (Mercer

et al., 2011). This later led to the identification of the mechanisms by which transcription factors may regulate each other, thereby acting as a switch between lineages. Finally, this paradigm has been further refined. Instead of emphasising individual key transcription factors, recent work suggests that transcription factors act as part of a network, specific for different stages of differentiation (Gottgens, 2015).

Transcription factors roles are dependent on the stage of haematopoiesis

Transcription factors may have a specific role in differentiation into particular lineages but this role may subsequently differ at other stages of haematopoiesis. For example, it was shown that mice which develop from a chimeric mix of *Scl*+/+, *Scl*+/- and *Scl*-/- ES cells show a failure of *Scl* knockout cells to contribute to haematopoiesis as HSCs fail to develop in the absence of *Scl* (Porcher et al., 1996). However conditional knockout of *Scl* after embryonic development show that lymphopoiesis and myelopoiesis is possible but erythroid and megakaryocyte differentiation remain compromised (Mikkola et al., 2003).

Despite the lineage defining roles of transcription factors they also play an important role at multiple stages of differentiation. *Cebpa* is absolutely required for the development of eosinophils and neutrophils (Zhang et al., 1997a). However, the *Cebpa* knockout also uncovered a role of this factor in adult HSC function (Ye et al., 2013). The inducible knockout of *Cebpa* in mice resulted in a pronounced increase in HSC numbers due to an increase in proliferation, with a transcriptional program that was more similar to foetal HSCs.

Transcription factors repress alternate lineage fates to enable lineage commitment

One route by which transcription factors may impart lineage specificity is through repression of alternative cell fates. One example is that of *Pax5.* In *Pax5-/-* mice pro-B cells fail to develop into mature B cells. This capacity is rescued by retroviral expression of *Pax5*. *Pax5-/-* pro B cells remain capable of

differentiating into granulocytes, neutrophils and other lineages when exposed to appropriate cytokines, because they aberrantly express genes from other non-lymphoid lineages (Nutt et al., 1999). Therefore, it appears that PAX5 is required to repress myeloid specific genes whilst the subsequent expression of other transcription factors such as E2A and EBF can then drive expression of B cell specific genes in the B lineage committed cell.

GMPs develop into either granulocytic neutrophils or monocytes/macrophages. This decision rests in part with *Gfi-1*. This transcription factor is important for the development of terminally differentiated neutrophils: rescue of GFI1 in atypical *Gfi-1-/-* progenitors result in differentiation of neutrophils, presumably through suppression of the alternative monocytic fate (Hock et al., 2003). In contrast, *Pu.1*-/- haematopoietic progenitors could be rescued through retroviral expression of *Pu.1,* which results in macrophage but not neutrophil development (DeKoter and Singh, 2000). Furthermore, *Pu.1* haploinsufficiency results in preferential development of GMP into neutrophils at the expense of monocytic differentiation (Dahl et al., 2003). Mechanistically, there appears to be a direct physical association between Gfi-1 and PU.1 that prevents PU.1 mediated activation of target genes (Dahl et al., 2006), thereby affecting the balance between neutrophil versus monocyte development.

Lineage re-programming properties of transcription factors

One of the most potent lines of evidence that transcription factors are lineage instructive is their ability after ectopic expression to trans-differentiate cells and enable lineage switching. For example, ectopic expression of GATA-1 in CLPs enforces their development into megakaryocytes and erythrocytes. However, the effect of GATA-1 still depends on which cell type it is expressed in: when GATA-1 was transduced into pro-B cells and GMP, it appeared to induce apoptosis (Iwasaki et al., 2003).

C/EBP transcription factors are capable of trans-differentiating lymphoid cells into myeloid cells (Xie et al., 2004, Laiosa et al., 2006, van Oevelen et al.,

2015). Xie et al showed that primary B cell precursors are capable of conversion into macrophages after the retroviral expression of C/EBPα (Xie et al., 2004). Although fully mature B cells could also undergo trans-differentiation, the conversion rate was lower. This conversion process lasted 3-4 days, with an intermediate stage whereby both lymphoid and myeloid surface markers were expressed. The efficiency of reprogramming was proportional to the level of transcription factor expression. These mature macrophages were capable of phagocytosis, but retained evidence of immunoglobulin gene rearrangement. C/EBPβ was also capable of reprogramming but differed from C/EBPα in the longevity of re-programmed cells. Although C/EBPα and PU.1 acted synergistically in the ability to reprogram lymphocytes, PU.1 alone was incapable of reprogramming lymphocytes, but is required for the upregulation of Mac-1 by C/EBPα.

Typically, these trans-differentiation experiments involve the use of CLP as targets suggesting that this cell type has a particularly plastic epigenome: interestingly, the ability for transcription factors to switch myeloid progenitors to a lymphoid lineage has not been shown (Iwasaki and Akashi, 2007).

Transcription factors act in complexes which include chromatin re-modellers

Transcription factors form complexes with other transcription factors. This allows greater potential controlling different stages of development, by increasing the range of regulatory elements that can be bound. For example, *Tal1* knockout mice have demonstrated an important role for SCL/TAL1 in HSC and megakaryocyte/erythrocyte development. However, HSC development remains preserved even when SCL/TAL1 DNA binding capacity is lost, but SCL/TAL1 DNA binding capacity remains critical for megakaryocyte/erythrocyte development (Porcher et al., 1999). This suggest that in HSC development the predominant role of SCL is in stabilising a transcription factor complex, without necessarily binding DNA, whilst in megakaryocyte/erythrocytes regulatory elements that are directly bound by SCL/TAL1 are essential (Tijssen et al., 2011).

Transcription factors may also have synergistic effects with other co-operating partners in a complex. ChIP-seq of ten transcription factors in the HPC-7 cell line revealed novel interactions and transcription factor complexes. One example was a previously uncharacterised interaction between RUNX1 and GATA2/SCL. This was suggested by computational analyses showing clustering of ChIP-seq binding sites and their physical interaction was confirmed by co-immunoprecipitation of the transcription factors. Furthermore, although single heterozygous *Runx1* or *Gata2 +/-* mice were viable, compound heterozygous mice were not (Wilson et al., 2010). The authors suggested that this may be due to the fact that both GATA2/SCL and RUNX1 have a role in regulating the proliferative capacity of HSCs: *Runx1* knockout in adult mice reduced the number of long-term HSC and increased numbers of immature progenitors (Growney et al., 2005) whilst high *Gata2* levels increase quiescence in HSCs (Tipping et al., 2009).

Finally, transcription factor complexes include other co-activators, repressors and chromatin modifiers to directly affect transcriptional activity. One example is the interaction between EKLF and BRG-1. Gene targeting studies in mice revealed that EKLF is required for definitive haematopoiesis. EKLF forms an interaction with BRG1 which is part of the SWI/SNF complex (Brown et al., 2002). SWI/SNF complexes have nucleosome remodelling function and have the capacity to increase chromatin accessibility to allow gene transcription. *Brg1* mutants, which are unable to remodel chromatin result in the same phenotype as seen in EKLF knockout mice (Bultman et al., 2005).

Identification of transcription factor networks which regulate haematopoiesis

As described above, much of the early work in studying the importance of transcription factors in haematopoiesis has focused on the effects of the knockout of single genes. In more recent years, there has been an emphasis on how these transcription factors interact with each other, as this ultimately is responsible for the cellular identity. One way of understanding such a gene regulatory network is through the identification of cis-regulatory modules that

control the transcriptional activity of neighbouring genes. These cis-regulatory modules comprise of promoters/enhancers/insulators, which are bound by different sets of transcription factors. These sub circuits can then interact with each other through their output to build up a larger gene regulatory network (Pimanda and Gottgens, 2010). A cis-regulatory module that has been identified in HSCs consists of *Gata2-3, Fli1+12*, and *Scl+19* enhancers. They are active in HSCs and are bound by these transcription factors themselves, forming an auto-regulatory sub circuit (Pimanda et al., 2007). A key paper in this field (Novershtern et al., 2011) used gene expression analysis of 38 distinct cell populations of the adult haematopoietic hierarchy. Using the principle that co-regulated genes form part of the same sub-circuit, the authors combined this gene expression data with motif analyses from cis-regulatory modules to identify novel transcriptional networks for different haematopoietic lineages. They validated this data through ChIP-seq of four transcription factors in HSC and by shRNA against predicted members of these transcription factor networks in *in vitro* erythroid and myeloid differentiation of primary CD34+ cells.

Using this paradigm, ChIP-seq data from 16 transcription factors at different stages of an *in vitro* ES cell differentiation has been used to construct gene regulatory networks that define cellular identify at different stages of haematopoietic development (Goode et al., 2016) . The association of different transcription factor networks with different stages of differentiation provides strong evidence of its importance in regulating this developmental pathway.

1.2 Transcriptional regulation in eukaryotic cells

1.2.1 RNA polymerase and core transcriptional machinery

Central to the transcriptional process is the binding of RNA polymerase II (RNAPII) to the core promoter and subsequent initiation and production of the RNA transcript. The initial assembly pre-initiation complex requires the recruitment of RNAPII with general transcription factors. Promoter clearance takes place as the complex begins the process of RNA transcription and results in RNAPII phosphorylation. RNA elongation requires separation from the

general transcription factor and recruitment of other proteins involved in RNA processing. Elongation continues until the whole of the gene is transcribed. Finally, termination of the process releases the elongation complex to allow the components to bind the promoter again to repeat the process (Jonkers and Lis, 2015).

The process of transcription encounters many obstacles, including other components of the DNA replication and damage response pathways. The RNAPII complex also has to traverse through nucleosome bound DNA (Bondarenko et al., 2006, Kireeva et al., 2005). A number of models have been posited, to explain how this takes place. It has been suggested that nucleosomes encountered by RNAPII may be displaced and retained by chaperones until the RNAPII complex has traversed through (Belotserkovskaya et al., 2003, Kireeva et al., 2002). Another model suggests that histone acetylases increase chromatin accessibility as RNAPII progresses (Guermah et al., 2006, Carey et al., 2006). When the elongation process has completed, deacetylases return the histone to its former state (Selth et al., 2010) .

1.2.2 Structure and modification of nucleosomes

The nucleosome is the core subunit of chromatin and forms the basis of the chromatin fibre (figure 1-2). The nucleosome itself is a DNA and histone complex. It is comprised of a core formed from a pair of H3 and H4 histones contacting a 70bp coil of DNA. A pair of H2A and H2B is formed above and below this, each conjoined to a further 40bp of DNA, thus the basic histone octamer comprises of 147bp DNA (Luger et al., 1997). After separation by further linker regions these nucleosomes are spaced 200bp apart. Secondary structures of 30nM fibres are then formed. There is evidence that subsequently, these secondary structures can be folded into much higher levels of compaction, with subsequent de-compaction upon gene activation in the region (Hu et al., 2009, Janicki et al., 2004).

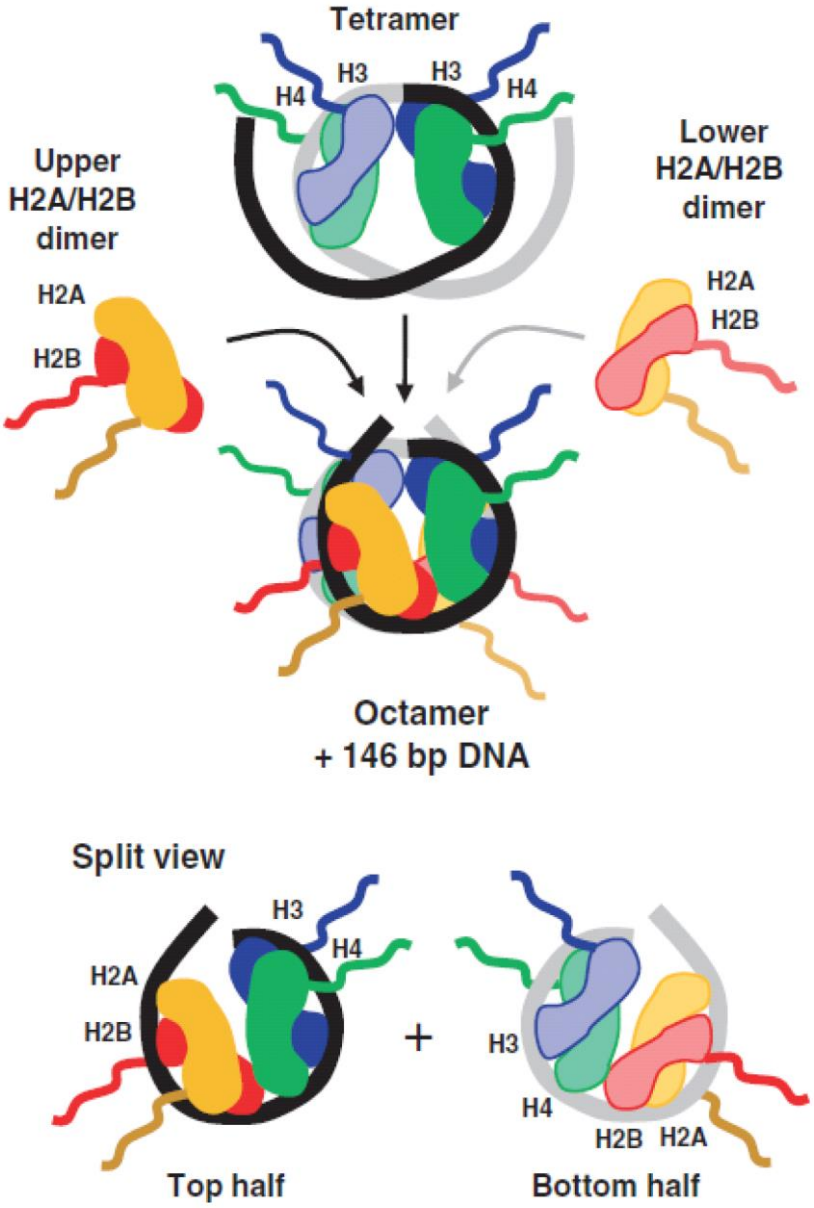


Figure 1-2 Histones and DNA interact to form nucleosomes.
A tetramer formed of two pairs of H3/H4 histones are paired with 70bp coil of DNA. This is subsequently joined by two pairs of H2A/H2B histones each with 40bp DNA to form the basic nucleosome octameric structure (Figure reproduced from publication (Cockerill, 2011))

Post-translational histone modifications

The resolution of the crystal structure of the nucleosome demonstrated that the N-terminal of histone tails protruded out of the core octamer subunit (Luger et

al., 1997) (figure 1-2). The majority of post-translational histone modifications have been characterised on these tails, but they have also been found in the histone globular core. Histone acetylation and methylation were the original modifications to be identified and were recognised to be post translational modifications with the capacity to influence RNA synthesis (Allfrey et al., 1964). The acetylation of histone lysines are accomplished by histone acetylases (HATs), and subsequently removed by histone deacetylases (HDACs). These histone modifiers often have little specificity in themselves, and act in complexes with other enzymes and transcription factors. Other than lysines, a number of other residues that can be modified have been characterised, including serine and arginine. Furthermore, an increasing number of modifications have been discovered, including phosphorylation, ubiquitination and sumolyation (Bannister and Kouzarides, 2011).

Function of histone modifications

A number of hypotheses have been suggested as to the mechanism by which histone modifications can affect RNA transcription. It has been postulated that the neutralisation of the positive charge in the histone N terminal weakens the interaction between the histone and the bound DNA, thereby making it easier to displace them (Bannister and Kouzarides, 2011). Consistent with this hypothesis, is that histones are acetylated prior to displacement at active promoter regions (Reinke and Hörz, 2003). Histone modifications may also affect the higher order structure of chromatin. One example is the acetylation of H4 at lysine 16 (H4K16). The incorporation of a chemically ligated H4K16ac into a nucleosome array prevented the formation of the 30nM chromatin fibre and inhibited interaction between fibres (Shogren-Knaak et al., 2006).

Histone modifications can also be read by other proteins, through which complexes with other functions can be recruited. As such it has been suggested that combinations of histone modifications act as a “code” and can be “read” by other proteins to result in a range of downstream effects (Turner, 2002, Turner, 2012, Strahl and Allis, 2000). For example, acetylated lysines can be bound by bromodomains whilst methylated lysines can be bound by

chromodomains or PHD regions (Li et al., 2007). Proteins with the same domain can recognise different residues despite the presence of the same modification. For example, heterochromatin gene repression is dependent on HP1. Although HP1 has a chromodomain, it specifically recognises tri-methylated H3 lysine 9 but not tri-methylated H3 lysine 4 (Bannister et al., 2001).

Transcription factors and chromatin remodellers modify nucleosome structure

"Pioneer factors" describe proteins that bind to DNA through compact chromatin (Zaret et al., 2016). Their binding precedes the subsequent binding of other larger protein complexes. Therefore pioneer factors have to have the ability to decondense chromatin and thereby increase its accessibility (Zaret and Carroll, 2011). One possible mechanism by which this may occur is illustrated by the FOXA transcription factors which possess DNA sequence specific binding domains but also has another domain that can contact neighbouring core histones and is necessary for the chromatin modifying capacity of the transcription factor (Cirillo et al., 2002). It is suggested that this ability to bind both core histones and the neighbouring DNA disrupts the nucleosomal structure.

Chromatin remodellers are a class of ATP dependent enzymes that can persistently displace nucleosomes (Owen-Hughes et al., 1996). A number of mechanisms have been suggested to explain the disruption of nucleosomes from the DNA, but the key requirement is for the hydrolysis of ATP to provide the energy required for this reaction to occur. These remodelling complexes do not work alone, and will interact with other DNA binding proteins such as transcription factor complexes and RNAPII. They also require histone chaperones to bind evicted histones and prevent their subsequent re-assembly (Schwabish and Struhl, 2006).

1.2.3 DNaseI hypersensitive sites

The observation that active genes have increased sensitivity to nucleases, such as DNaseI, was made by noting that β globin gene fragments could be isolated through the nuclease treatment of nuclei from erythrocytes but not from fibroblasts (Weintraub and Groudine, 1976). It was further observed that sites postulated to be involved in the regulation of the adjacent gene encoding heat shock protein in Drosophila was DNaseI sensitive (Wu, 1980). DNaseI hypersensitivity sites (DHSs) can represent tissue specific enhancers, consistent with their importance in the specific control of gene expression (Cockerill et al., 1999). Fox example, DHS mapping has been use to locate the regulatory elements that control globin gene loci. By using a "mini-locus" that contain the human β-globin gene and the flanking DHSs, investigators have been able to stably reproduce, in transgenic mice, the expression of the gene at levels observed in humans (Grosveld et al., 1987). DHS can also represent other elements that include promoters, insulators, and locus control regions (Cockerill, 2011) .

DNaseI hypersensitive sites from DNase-seq and their association with other epigenetic features

DHSs have been mapped genome wide through the use of tiling arrays and subsequently by massively paralleled sequencing technology (Boyle et al., 2008). By coupling DNaseI digestion to next generation sequencing (NGS) (e.g. Illumina platform) authors were able to produce base pair resolution mapping of DHSs in CD4+ T cells. They showed that vast majority of DHSs are distal to the promoters of genes. From their data nearly all highly expressed genes have a DHS, but a presence of a DHS did not necessarily mean that the proximal gene is expressed.

The analysis of DHSs with corresponding histone modifications from ChIP-seq confirms that the DHS peak correspond to a depletion of nucleosomes. However, histone modification enrichment was largely dependent on whether the nearest gene was transcribed. For example, average profiles for tri-methylated H3 lysine 4 showed that it was generally present at DHS but the

level of enrichment was dependent on the transcriptional activity of the gene nearest to the DHS. In contrast, trimethyl H3 lysine 9 and H3 lysine 27, as expected from their role in gene repression, were absent from DHS profiles. In comparison to predictions based on histone modifications alone, using ENCODE data, DHSs outperformed histone modification states in predicting enhancer activity (Kwasnieski et al., 2014), however not all DHSs are enhancers (Dogan et al., 2015).

Across a number of cell lines increased CpG methylation is almost uniformly associated with a lack of DHS (Thurman et al., 2012). However, when looking at all CpG containing DHS, a simple linear regression showed that only 20% of these sites had a significant association between methylation and accessibility.

Analysis of sequences from distal DHS in CD4+ T cells (Boyle et al., 2008) showed that motifs relevant to the cell type could be recognised. Nearly all (94.4%) transcription factor binding sites catalogued through ChIP-seq, by the ENCODE project, reside in DHSs (Thurman et al., 2012). The small percentage of transcription factors that do not, are associated with heterochromatin maintenance, such as SETDB1.

DNaseI footprinting identified regions occupied by transcription factors

DNaseI footprinting is based on the observation that DNA is protected from cleavage by nucleases by the presence of DNA-bound proteins that interact with the bases in the DNA such as transcription factors (Dynan and Tjian, 1983). This principle can be used to map transcription factor binding across the genome, although this may not be possible for highly dynamic transcription factor-DNA interactions (Vierstra and Stamatoyannopoulos, 2016, Sung et al., 2016). The use of this principle at a genome wide level critically depends on the density of nuclease cleavages at these DHSs. For the purpose of genome wide DNaseI footprinting, in the ENCODE project, an average of 273 million reads per cell type were sequenced (Neph et al., 2012). In this paper, they showed that the probability of detecting a footprint was proportional to the ChIP-seq signal for that transcription factor. They also showed that heterozygous

single nucleotide variants (SNVs) in binding motifs can result in the loss of a transcription factor footprint. Differences in chromatin accessibility between alleles of DNA were dependent on the presence or absence of these SNVs. Unsurprisingly, they were also able to show that DNaseI footprinting between different cell types were able to identify different patterns of transcription factor footprints.

DNA cleavage pattern around a footprinted protein and the adjacent nucleosome can be used to optimise footprinting analysis pathways. A recent algorithm takes advantage of this to improve transcription factor footprinting predictions from high read depth DNase-seq data (Piper et al., 2013). DNaseI cleavages within a footprinted location can be compared with the number of DNaseI cleavages within the surrounding DHS. DNA bound transcription factors lying adjacent to nucleosomes offer protection to the DNA from the nucleases. This results in longer DNA fragments than is otherwise found within the DHS (figure 1.3). There is a reduced probability of identifying a fragment that aligns downstream from the transcription factor complex because the second cut will then have to occur outside the DHS in the nucleosome region. Information from each strand is assessed independently by making use of the sequencing from both the positive and negative strands (figure 1-3).

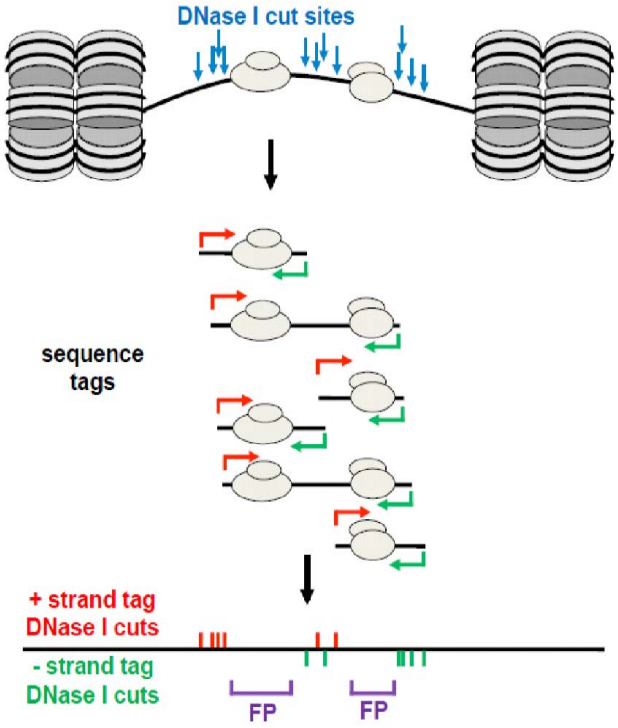


Figure 1-3 DNaseI Hypersensitivity site mapping and transcription factor footprinting.
DNaseI preferentially cleaves DNA at nucleosome free regions. Transcription factor binding sites are protected from DNAseI cleavage relative to the surrounding DHS. Fragments upstream of transcription factor binding sites aligning to the + strand (red), whilst fragments downstream of these binding sites align to the - strand (green) (From Piper et al 2013).

1.2.4 Regulation of transcription through DNA modifications

The best characterised DNA modification is methylation at position 5 of cytosine (5mC) and this is accomplished by DNMT3a and DNMT3b (Okano et al., 1999). 5mC is a highly conserved modification found in many animal, plant and fungi (Wu and Zhang, 2014). This modification can be subsequently maintained by DNMT1 and its partner UHRF1. DNMT1 recognises methylated CpG islands on one strand and subsequently methylate the cytosine on the newly replicated strand (Fatemi et al., 2001, Hermann et al., 2004, Bostick et al., 2007). This ability to maintain 5mC allows the transmission of epigenetic information to

daughter cells and is crucial for normal development of many cell types. Correct maintenance of 5mC is also important for genomic stability: knockout of DNMT1 in cell lines result in mitotic catastrophe (Chen et al., 2007).

5mC like other epigenetic marks can be reversed. Two major routes of removing 5mC have been found, both of which converge on DNA damage pathways of repair. The two routes of 5mC removal differ from whether there is involvement of the TET family of proteins. The TET enzymes catalyse the progressive oxidation of 5mC into subsequent forms such as 5-hydroxymethylcytosine, 5 formylcytosine and 5-carboxylcytosine. These latter two species can be subsequently removed by base-excision repair. α-ketoglutarate is used as a co-factor in the oxidation of 5mC and is produced through the tricarboxylic acid cycle (TCA) by the enzyme Isocitrate dehydrogenase (IDH) 1 and 2 (Wu and Zhang, 2014). Both TET1/2 and IDH1/2 can be affected by mutations seen in acute myeloid leukaemia (AML).

Like other epigenetic modifications, 5mC can be read by other enzymes and thus can interact with other epigenetic pathways (Zhu et al., 2016). For example the enzyme MeCP2 is part of a family of proteins with a methyl-cytosine binding domain. MeCP2 acts as a transcriptional repressor and binds highly methylated promoters (Nan et al., 1997). This transcriptional repression is in part due to the recruitment of histone deacetylates by MeCP2 to these methylated CpG (Nan et al., 1998).

1.3 . Deregulation of transcriptional regulation in acute myeloid leukaemia

AML is a disease defined by the presence of at least 20% myeloid blasts, which represent early myeloid progenitor cells which are blocked in their differentiation pathway, in the peripheral blood or bone marrow aspirate. There is a worldwide incidence of approximately 3 cases per 100000 population per year (Swerdlow, 2008).

AML is the most common acute leukaemia in adults and despite improvements in supportive care, outcomes for these patients remains poor. Less than 60% of younger adults achieve longer term survival (Burnett et al., 2015) whilst those aged over 60, unfit for intensive chemotherapy one-year survival are less than 30% (Dennis et al., 2015).

1.3.1 Mutations found in acute myeloid leukaemia

A striking feature of AML is the clinical heterogeneity that underlies this diagnosis. This can be explained by the presence of different recurrent mutations that are associated with AML. For example, a patient with an AML expressing the fusion gene CBFB–MYH11 as a result of inv(16) translocation has over 60% chance of overall survival, in comparison to patients with monosomy 5 cytogenetic abnormalities who have a less than 10% chance of overall survival (Grimwade et al., 2010). This can be subsequently refined by new discoveries of driver mutations and re-classification by subgroups based on these findings (Papaemmanuil et al., 2016, Cancer Genome Atlas Research, 2013). Some of these findings are summarised in table 1-1. Newly discovered patterns of cooperativity and exclusivity may provide new avenues of research underlying the mechanisms behind the mutations. For example, the co-occurrence between NPM1c and FLT3-ITD mutations has been recognised for a long time and has been reproduced in a mouse model (Mupo et al., 2013). DNMT3a mutations and FLT3-ITD have been also more recently characterised (Meyer et al., 2016, Yang et al., 2016). However, the newly characterised group of patients with NPM1c-DNMT3a-FLT3-ITD mutations, which confers a particularly poor prognosis, (Papaemmanuil et al., 2016) has yet to be studied. An example of exclusivity, suggesting that mutations work in the same pathway is that of the TET and IDH mutations (Weissmann et al., 2012, Figueroa et al., 2010a). Here TET and IDH mutations are mutually exclusive. IDH mutations result in an overproduction of 2-hydroxygluatarate, depleting 2-ketoglutarate which is normally required for the oxidative capacity of TET enzymes. TET loss of function mutations thereby mimics gain of function mutations in IDH.

Table 1-1 Mutations in acute myeloid leukaemia

Class of mutations	Examples
Transcription factor	*RUNX1; C/EBPα; CBFB-MYH11; RUNX1-RUNX1T1; WT1;* translocations involving *MECOM* e.g. inv(3)
DNA methylation	*TET2; IDH1/2, DNMT3A*
Epigenetic modifiers	*MLL; ASXL1; EZH2*
Cohesin	*RAD21, SMC3, SMC1A, and STAG2*
Spliceosome	*SRSF2; U2AF1; SF3B1*
Signalling	*FLT3; KIT; NRAS; KRAS*
Others	Complex karyotypes; monosomy 5 and 7; *NPM1*

Even in newly discovered classes of mutations, where there are other functions of the mutated protein, the pathological effect of the mutation is that of transcriptional dysregulation. For example the spliceosome mutations affect mRNA splicing in the target cell, however, through overexpression and rescue experiments, the block in differentiation seen in these cells can be shown to be a result of the polycomb protein EZH2 downregulation (Kim et al., 2015). Whilst mutations in cohesin complex, involved in many cellular processes including DNA damage response, principally results in changes in chromatin accessibility for transcription factors such as ERG and GATA2 (Mazumdar et al., 2015).

Finally, the ability to describe prognostic gene expression and methylation profiles based on these classifiers confirm the view that these diseases are distinct biological entities driven by deregulated transcriptional profiles (Cancer Genome Atlas Research, 2013, Valk et al., 2004, Figueroa et al., 2010b).

Pre leukaemic clones and clonal heterogeneity in acute myeloid leukaemia

Cancers have long been defined by the uncontrolled growth of a clonal population of cells. However, recent use of NGS has increased sensitivity of mutation detection and through the use of variant allele frequency, has allowed detection of the prevalence of these mutations across the bulk population of cells that have been sequenced. These studies suggest that a significant proportion of AML consist of 2 or more subclones (Cancer Genome Atlas Research, 2013). One significance of this is the discovery that leukaemia relapse can occur through the accumulation of further mutations in a subclone (Ding et al., 2012).

This model also suggests the possibility of the existence of pre-leukaemic clones, of which one eventually dominants haematopoiesis through the acquisition of further mutations. One study showed that the accumulation of mutations in HSCs from otherwise healthy individuals are a function of age (Welch et al., 2012), which is in keeping with the predominance of AML in older adults. Subsequently, a number of targeted sequencing projects in large cohorts have identified recurrent mutations in *DNMT3A*¸ *ASXL1* and *TET2* in otherwise healthy individuals (Jaiswal et al., 2014, Genovese et al., 2014). This is in keeping with the increased self-renewal of *DNMT3A* knockout murine HSC (Challen et al., 2012) which confers upon these HSCs a selective advantage. This is also seen in engraftment of human pre-leukaemic DNMT3A mutant HSC, from patients in remission from AML, in xenotransplantation models, where they have a repopulating advantage over HSC without a mutation in DNMT3A (Shlush et al., 2014). The presence of mutations in HSCs is not a new idea as this has been previously demonstrated for both BCR-ABL and RUNX1-RUNX1T1, with the detection of these transcripts in both myeloid and lymphoid cells in patients in remission from their disease (Miyamoto et al., 2000, Passegue et al., 2003).

Understanding the subclonal architecture has many implications. One is the possibility that therapies affecting a clone with one mutation may result in the

outgrowth of a subclone with another mutation unaffected by this therapy (Wong et al., 2015) (Greaves and Maley, 2012). A second is the concern that minimal residual disease monitoring of a secondary mutation, e.g. *NPM1c*, may miss a relapse from another mutation in a pre-leukaemic clone. However, this has not been borne out in clinical studies (Ivey et al., 2016) which suggests that although pre-leukaemic clones are present, the path to the development of a full AML is stereotyped, due to cooperativity required between mutations.

Leukaemic stem cells in acute myeloid leukaemia

The concept of leukaemic stem cells (LSC) (or leukaemia initiating cell) in AML proposes that only a subpopulation of cells within the bulk leukaemic blast population has leukaemia initiating potential, which is defined as the capacity to initiate leukaemia in a xenotransplantation model. According to this model, to prevent relapse after remission induction therapy, it is vital to eradicate this population of cells. A pioneering study showed that the frequency of these cells were approximately 1 in $1x10^6$ CD34+CD38- AML cells, and was defined as leukaemia initiating through their ability to transplant a leukaemia into an immunodeficient (SCID) mice (Lapidot et al., 1994). Subsequent further characterisation by the group demonstrated that a single LSC had the potential to generate millions of "differentiated" blasts of various immunophenotypes, and could serially repopulate irradiated mice confirming they had self-renewal capacity (Bonnet and Dick, 1997). From this the authors proposed that that the normal CD34+CD38- HSCs are the target for transformation in AML.

However, subsequent studies have sought to further refine characteristics of LSC to enable their separation from normal HSC, through identification of cell surface markers. These have included CD47, TIM3 (Kikushige et al., 2011) and CD32/CD25 (Saito et al., 2010). CD47 expression transiently expressed on normal HSC is constitutively expressed on LSC to prevent phagocytosis from circulating macrophages (Jaiswal et al., 2009). Furthermore, studies also discovered that some LIC were not CD34+, but resided in a CD34- population, typically these are NPM1 mutant AML (Taussig et al., 2010, Quek et al., 2016). Taken together, this suggests that the immunophenotype of LSC may be

heterogeneous (perhaps a reflection of the underlying genetic heterogeneity) and is dependent on the target cell of transformation but may also reflect the inherent differences between models of immunodeficient mice used which may display differences in engraftment potential. Given that the "gold-standard" for LSC remain that of engraftment in mice, the lack of a suitable mouse recipient has hindered research into LSCs in certain leukaemia such as t(15;17) or t(8;21).

Although the evidence suggests that HSCs are cells in which mutations can occur, further studies are now more consistent with the model that the LSC arises from a more committed progenitor. This include evidence from CML where the GMP population were expanded in patients in blast crisis and these cells have evidence of increased self-renewal properties (Jamieson et al., 2004) and in AML where the CD34+ LSC is more in keeping with LMPP expression profile (Goardon et al., 2011) and the CD34- LSC more in keeping with the GMP profile (Quek et al., 2016). This suggests that committed progenitors may aberrantly acquire stem cell characteristics such as self-renewal as part of leukaemic transformation. The proof-of-principle came from studies of MLL fusion leukaemias where investigators showed the ability to transform committed GMPs into LSC through overexpression of MLL-AF9 (Krivtsov et al., 2006).

To understand the mechanism behind the self-renewal properties of LSCs investigators have used gene expression profiling of HSCs and compared them to LSCs to identify an aberrant transcriptional profile (Eppert et al., 2011) and in keeping with another study (Gentles et al., 2010) were able to show that this LSC gene expression pattern can be seen in patients and had prognostic significance. This lends credence to the LSC hypothesis, suggesting that this model is not only an artefact of xenotransplantation. However, the degree of overlap in genes identified by these studies was small and although in vivo testing for LSC were performed, the frequency of true leukaemia initiating cells in the population identified as "LSC" used for gene expression profiling was extremely low (Eppert et al., 2011).

1.3.2 Targeting transcriptional deregulation in acute myeloid leukaemia

The standard treatment for AML has been a combination of anthracycline and cytarabine for many decades. This course of treatment is intensive and restricted to patients fit enough to withstand the side effects. Even, in this setting relapse or treatment refractory disease remains the most common cause for treatment failure. Therefore, targeted therapy (table 1-2) holds promise in providing novel, more tolerable clinical options (Shafer and Grant, 2016). The practical considerations of this strategy are many fold and include rapid turn-around of mutational analyses in a clinical setting in order to identify patients who might benefit from these drugs. Furthermore, even if proved successful the expense of funding novel agents, even in affluent healthcare systems, may prove prohibitive to many patients.

Table 1-2 Examples of novel targeted epigenetic therapies in clinical trials (CR= complete remission, CRi= complete remission with incomplete count recovery)

Name of agent	**Target**	**Clinical responses**	**Clinical trial identifier**
AG221	Mutant IDH2	11/32 CR or CRi	Phase III (NCT02577406)
OTX015	BRD2/3/4	2/33 CR or CRi-ongoing study	Phase I (NCT01713582)
GSK2879552	LSD1	No results available	Phase I (NCT02177812)
EPZ-5676	DOT1L	1/28 CR-ongoing study	Phase I (NCT01684150)

Targeting regulators of DNA methylation

Azacitidine and decitabine are cytosine analogues that can be incorporated into DNA and can inhibit DNMT1 leading to widespread hypomethylation (Kelly et al., 2010, Klco et al., 2013, Yan et al., 2012). It has been approved for the treatment of higher risk myelodysplastic syndrome through a Phase III clinical trial in this setting (Fenaux et al., 2009). Although there is evidence of a conventional cytotoxic effect at higher doses, another mechanism may be re-activation of endogenous retroviral elements through demethylation of 5mC, leading to stimulation of the host immune system (Roulois et al., 2015, Chiappinelli et al., 2015).

As discussed above, IDH gain of function mutations are seen in AML. Oral inhibitors of mutant IDH have seen good tolerability and impressive responses. Furthermore investigators have been able to use the aberrantly produced metabolite α-HG as a biomarker of response (Stein et al., 2014).

Targeting epigenetic readers and modifiers

Bromodomains confer the ability to bind to acetylated histones to the BET family of proteins such as BRD2/3/4, thus acting as an adaptor for the recruitment of members of the transcriptional elongation machinery. In AML cells with MLL translocation, knockdown of BRD4 by shRNA or inhibition by small molecules resulted in cell death (Zuber et al., 2011, Dawson et al., 2011). The effects of BET inhibition include the reduction in transcription of genes such as *MYC* and *BCL2*. BET bromodomain inhibition may also function through inhibiting the recruitment of key transcription factors, to the regulation of important downstream genes (Roe et al., 2015).

LSD1 also known as KDM1A is a histone lysine demethylase of both dimethylated histone 3 lysine 9 (thereby activating target genes) and also dimethylated histone 3 lysine 4 (thereby repressing target genes). It appears to have a crucial role in maintaining the leukaemic stem cell phenotype of MLL leukaemia models (Harris et al., 2012). An oral inhibitor of LSD1, GSK2879552, is currently undergoing trials in treatment of relapsed refractory AML.

The histone methyltransferase DOT1L is a binding partner of MLL fusion proteins and methylates histone 3 lysine 79, increasing gene expression of targets such as *MEIS1* and the *HOXA* cluster. Inhibition of DOT1L in a conditional knockout system resulted in differentiation of AML cells expressing MLL-AF9 (Bernt et al., 2011). A DOT1L inhibitor, EPZ-5676, reproduced these findings in a pre-clinical model (Daigle et al., 2013) and preliminary results from a Phase I clinical trial shows some signs of clinical activity.

Targeting transcription factor complexes

A recent, example of targeting transcription factor interactions is the use of a small molecule inhibitor to bind disrupt the interaction between RUNX1 and the fusion protein CBFβ-SMMHC, a product of the chromosomal translocation inv(16). AI-10-46 specifically affected AML cells that had the inv(16) abnormality. By disrupting the association between CBFβ-SMMHC to RUNX1, normal RUNX1 binding to native CBFβ was restored. This allowed RUNX1 to bind DNA normally, and activate its target genes (Illendula et al., 2015).

1.4 Role of RUNX1 in normal haematopoiesis

RUNX1 is a master regulator of haematopoiesis and is essential for both early development and emergence of haematopoietic stem cells as well as subsequent myeloid lineages and lymphopoeisis (Speck and Gilliland, 2002). RUNX1 is a member of the core-binding factor transcription factor family and forms a heterodimer with core-binding factor Beta (CBFβ), which is important for RUNX1 activity. Tissue specific expression of *Runx1*, *Runx2* and *Runx3* in mice suggests a role in developmental regulation. DNA interaction is mediated by the evolutionarily conserved Runt domain (Daga et al., 1992) which can bind to the consensus DNA sequence TGTGGT (Thornell et al., 1988, Speck and Baltimore, 1987).

1.4.1 Role of Runx1 at the onset of haematopoiesis

Runx1 is essential for development of definitive HSCs in the AGM. When *Runx1* is knocked out in mice by homologous recombination, they exhibit embryonic lethality at E12.5 due to a failure of fetal liver haematopoiesis (Okuda et al., 1996). The embryos are anaemic, although primitive erythrocytes are present and death is a result of haemorrhage secondary to severe thrombocytopaenia. Transplant of *Runx1-/-* embryonic stem cells into *Runx1 +/-* blastocysts created chimeric mice, *Runx1-/-* cells failed to develop into haematopoietic tissue confirming that the defect resides in the mutant cells and not due to a defect with the HSC niche.

A *Runx1*-lacZ knock in embryo has been generated with exons 7 and 8 of *Runx1* replaced with *lacZ* sequences. The resultant fusion protein is non-functional as a *Runx1* allele. Thus, in a mouse hemizygous for a normal/*Runx1*-lacZ RUNX1 gene, haploinsufficient *Runx1* expressing cells can be isolated. It can be shown that all HSC express *Runx1* and that the prevalence of HSCs and their distribution in different compartments is sensitive to *Runx1* dosage (North et al., 2002). Through transplantation assays it can be demonstrated that all developing HSCs, found in the AGM, fetal liver and the vitelline uterine artery are *Runx1* positive. RUNX1 can be detected as early as E7.5 in mesenchymal cells located in the yolk sac.

1.4.2 Role of Runx1 in Adult Haematopoiesis

Given the embryonic lethality of mice with full *Runx1* knock-out alleles, a conditional knock out of *Runx1* was generated to examine the role of RUNX1 in adult haematopoiesis (Ichikawa et al., 2004, Growney et al., 2005, Putz et al., 2006). Mice, where exon 5 of the *Runx1* gene was flanked by *lox*P sites, were crossed with *Mx-cre* mice (Ichikawa et al., 2004). Homozygous *Runx1* deleted mice could be generated by injection of polyinosinic-polycytidylic acid. Although these mice produced mature lymphocytes, the proportion of lymphocytes with biallelic deletions in thymocytes and B cells were only 9% and 57%, respectively, suggesting their development was impaired. Erythrocytes and

neutrophil counts were unaltered, but platelet counts dropped substantially to up to a sixth of the initial numbers. A striking result was the presence of biallelic deleted *Runx1* HSC that were able to repopulate mice in a competitive repopulation assay. Thus, although *Runx1* is essential for their development, it is not required for the maintenance of HSCs. Although definitive lymphocyte development was absent, immature T cell precursors accumulated in the thymus. Again, normal myeloid precursors and morphologically normal neutrophils were present in expected numbers. Others were in agreement with this phenotype except, in addition, some have noted a mild myeloproliferative phenotype in adult mice with knockout of *Runx1* (Growney et al., 2005).

Due to the disrupted lymphopoiesis in adult mice with conditional *Runx1* knock out, the role of *Runx1* in lymphocyte development has been examined further. To specifically target early T cell progenitors, *Runx1-lox*P flanked homozygous mice were crossed with *Lck-Cre* mice (Taniuchi et al., 2002). T lymphocyte development involves the transition of double negative (DN) CD4-CD8- cell to double positive CD4+CD8+ cells, and finally to mature single positive (SP) CD4+ or CD8+ cells. Although analysis of thymocytes from these mice revealed 95% population of cells had the genotype *Runx1*-/-, substantial numbers of SP CD4+ (23%) or CD8+ (59%) cells were *Runx1+/+,* suggesting *Runx1* may have a role in thymic selection. However, given that a significant number of mature SP cells retained the *Runx1-/-* genotype, RUNX1 is not essential in maintaining CD4 or CD8 silencing. *Runx1* is also involved in later lineage decisions between Th1 and Th2 cells. *Gata3* is required for the differentiation of T lymphocytes into Th2 phenotype cells. *Runx1* overexpressing cells showed a skew to Th1 cells with the mechanism being an inhibition of *Gata3* expression. Indeed, the Th2 phenotype could be rescued through forced *Gata3* expression(Komine et al., 2003).

In humans, families have been identified that carry a Familial Platelet disorder gene with predisposition to acute myelogeneous leukaemia (FPD/AML). These pedigrees show an autosomal dominant pattern of inheritance. Patients with FPD/AML are thromobocytopaenic, with dysfunctional platelets and have an

increased risk of developing AML (Song et al., 1999). Linkage disequilibrium studies using genetic markers defined a 3.2Mb genetic region of interest on chromosome 21. Sanger sequencing of the candidate gene *Runx1* identified a number of mutations, all of which resulted in the loss of the transactivation domain, and/or point mutations resulting in a truncated form of the protein or disruption to the RUNT DNA binding domain. Bone marrow specimens from affected patients were used in megakaryocyte colony forming assays, which confirmed the presence of defective megakaryopoiesis. Although there was no evidence of the inappropriate expansion and maturation arrest of myeloid cells subsequent analysis of leukaemic cells from these patients showed retention of the wild-type allele in cells showing clonal chromosomal abnormalities.

In summary, *Runx1* is essential for the development but not the maintenance of HSC. It has a clear role in the development and further differentiation of T lymphocytes. Although B lymphocytes numbers were reduced in adult mice with conditional knock out of *Runx1* the mechanism of *Runx1* function in B lymphopoiesis is not fully elucidated. Patients with FPD/AML demonstrate impaired megakaryopoiesis: a phenotype which is reproduced in the murine conditional *Runx1* knockout model or mice with the same mutations as those seen in patients with FPD/AML (Matheny et al., 2007, Sun and Downing, 2004).

1.4.3 RUNX1 as a transcription factor

RUNX1, and the drosophila homologue Runt, binds a murine retrovirus enhancer element in a sequence specific manner (Thornell et al., 1988, Speck and Baltimore, 1987). RUNX1, also known as core-binding factor alpha (CBFα), functions as a heterodimer with core binding factor beta (CBFβ), which increases its DNA binding affinity by up to tenfold (Wang et al., 1996). CBFβ does not interact directly with DNA, but restricts the mobile regions of the Runt domain of RUNX1, resulting in a more stable conformation with DNA (Tahirov et al., 2001). Mice with homozygous deletions of CBFβ die from widespread haemorrhage with an identical phenotype to *Runx1* knock out counterparts (Wang et al., 1996).

Transcription factors cooperate with other partners to increase tissue specific DNA binding activity. RUNX1 in mammalian cells tends to associate with other haematopoietic activators such as the C/EBP or ETS family transcription factors, through adjacent motifs localised in enhancer regions of genes encoding markers of monocytic differentiation such as cytokine receptors (Zhang et al., 1994, Petrovick et al., 1998). For example, the tyrosine kinase receptor CSF-1R, is encoded by a gene whose expression is restricted to monocytes, and is essential for their proliferation and survival (Dai et al., 2002). A short -87 to-29 bp region in the promoter region encodes a binding sequence for PU.1 (a member of the ETS family of transcription factors), as well as RUNX1 and C/EBPα. Furthermore, there is direct physical interaction between C/EBPα, and RUNX1, as demonstrated by co-immunoprecipitation (Co-IP) and by a reduction in luciferase promoter activity when a construct is generated that increases the distance between these binding motifs. In these reporter assays basal levels of transcription generated by the heterodimer RUNX1 and CBFβ is 80 fold increased by the addition of C/EBPα. A further intronic regulatory element (FIRE) in this gene has also been identified (Himes et al., 2001). Here, an Ets/Runx1 site co-exists and through ChIP direct occupancy of RUNX1 can be demonstrated (Tagoh et al., 2002). RUNX1 binding sites often co-localise with ETS transcription factor binding sites (Gunther and Graves, 1994): the presence of ETS family transcription factors improves DNA binding ability of RUNX1 (Gu et al., 2000). Recent ChIP seq data have confirmed the co-localisation of ETS motifs with RUNX1 binding sites (Ptasinska et al., 2012, Ben-Ami et al., 2013).

Transcriptional co-activators that do not bind DNA directly act in complex with RUNX1 to modify histones to increase accessibility and recruit transcription initiating complexes. Two well-characterised co-activators are p300 and CREB binding protein (CBP). These are multifunctional adaptor proteins that bind a variety of transcription factors including AP-1 (Arias et al., 1994, Kamei et al., 1996) and c-Myb (Zor et al., 2002, Dai et al., 1996). P300 and CBP can interact with the basal transcription factor TFIIB (Kwok et al., 1994) and can acetylate histone H3 independently (Oelgeschlager et al., 1996) and via a partner protein,

P/CAF (Yang et al., 1996) to increase chromatin accessibility for other factors. A C-terminal region of RUNX1, lost in CBF fusion proteins, is essential for the physical interaction of p300 and CBP to RUNX1 (Kitabayashi et al., 1998). Luciferase activity reporter assays with the Myeloperoxidase (MPO) promoter, a gene important for neutrophil function, suggests both p300 and RUNX1 are required for its transcriptional activation. Other co-activating partners such as ALY and YAP function in similar cooperative fashion.

Outside of the myeloid system, RUNX1 has been shown to act in the context of transcriptional repression. As described above RUNX1 plays an important role in lymphocyte differentiation, in particular in silencing CD4 gene expression. Although direct evidence is lacking this may be due to the recruitment of the BAF chromatin remodelling complex (Chi et al., 2002). In the context of p21 gene regulation RUNX1 is also able to interact with the repressive chromatin modifiers, mSin3A and mSin3B, which exists in complexes associated with histone de-acetylases (Lutterbach and Hiebert, 2000).

1.5 Role of *EVI-1* in normal haematopoiesis

EVI-1 (also known as *MECOM* or *PRDM3*) is found on chromosome 3q26 and is a dual domain zinc finger transcription factor (seven at the N-terminal and three at the C-terminal). Evi-1 in itself is an essential regulator of self-renewal in haematopoietic stem cells (Goyama et al., 2008).

1.5.1 Role of EVI-1 in fetal haematopoiesis

Expression of *EVI-1* in the haematopoietic system is limited to CD34+ precursor cells (Gerhardt et al., 1997, Privitera et al., 1997), including HSCs (Phillips et al., 2000) and becomes down-regulated in maturing cells (Russell et al., 1994). However, EVI-1 is likely to have an important role outside of the haematopoietic system. Widespread abnormalities are seen in mice with lack of *Evi-1*: the lack of *Evi-1* is lethal at day 11 of embryonic development with widespread

abnormalities seen in the cranial system, somites and peripheral nervous system (Hoyt et al., 1997).

In murine *Evi-1* -/- embryos, the defect in haematopoiesis is seen in cells of the para-aortic-splanchno-pleural (P-Sp) which in wild-type embryos have a high *Evi-1* expression and show a decrease in frequency of HSCs isolated in the knock-out. *In vitro* stromal co-culture studies of these isolated cells demonstrate a reduced proliferative capacity and after adoptive transfer into a myeloablated murine host, these cells are incapable of establishing haematopoiesis at all (Yuasa et al., 2005). Gene expression analyses suggest that *Evi-1* knockout results in a substantial decrease in *Gata-1* and *Gata-2* mRNA levels (Yuasa et al., 2005). *Evi-1* knockout P-Sp HSCs transfected with a retrovirus that enforced the expression of *Evi-1* or *Gata-2* were able to rescue the function of these *Evi*-1 deficient HSCs. *Gata-2* expression is regulated through the use of two promoters: haematopoietic cells employ the IS promoter, whilst a 3' proximal IG promoter is used in other tissues (Pan et al., 2000). Yuasa et al (Yuasa et al., 2005) suggest that a region between 6-7kb upstream of this promoter contains a DNA binding motif for the first of the two zinc finger domains of EVI-1 located towards the N-terminus. When this regulatory element was cloned into a luciferase reporter construct, and co-transfected into a haematopoietic cell line EML with an *Evi-1* overexpression vector, this *Gata-2* regulatory element enhanced luciferase activity compared to other upstream regions of this gene. Yuasa et al (Yuasa et al., 2005) also demonstrate by chromatin immunoprecipitation-PCR (ChIP)-PCR in this cell line that EVI-1 binds this enhancer specifically.

1.5.2 Role of EVI-1 in adult haematopoiesis

Due to the embryonic lethality of *Evi-1* knockout a conditional knock-out model has been generated to study the role of *Evi*-1 in adult haematopoiesis (Goyama et al., 2008). This was accomplished by crossing mice with floxed *Evi-1* with Mx-Cre expressing mice. Mature myeloid cells and lymphocytes were

unaffected when Evi-1 was excised. However, the frequency of HSC was greatly reduced and *Evi-1*-/- HSCs lacked self-renewal and differentiation capacity as they did not contribute to mature haematopoietic cells when the mice were examined 4 weeks later. *Evi-1* +/- mice had a reduced number of HSCs and when these cells were transplanted with competitor wild-type HSCs, there was a reduced fraction of *Evi*-1+/- cells, indicating correct levels of *Evi-1* expression is important. This defect in HSC function can be rescued by transfecting an *Evi-1* expression vector (Kataoka et al., 2011).

IRES-GFP labelled Evi-1 transgenic mice were generated to enable accurate sorting of *Evi-1* expressing cells during development (Kataoka et al., 2011). The authors confirmed that *Evi-1* was only expressed in the HSC population with subsequent loss of *Evi-1* expression upon differentiation in vitro and in vivo studies. Furthermore, the expression of *Evi-1* is restricted to only the earliest long-term HSC which are naturally quiescent, possess self-renewal properties and have the most potent multi-lineage differentiation capacity.

In summary, Evi-1 has an essential role in the development and maintenance of long-term HSCs, without any substantial role in mature differentiated cells.

1.5.3 Features of EVI-1 as a transcription factor

Several studies have shown that EVI-1 can bind DNA in a sequence specific manner through ten zinc fingers located in two separate domains situated towards the C-terminal. From one study, a gel shift PCR method determined the consensus binding site for the first domain of seven zinc fingers as TGACAAGATAA (Perkins et al., 1991). Further work to extend this, using GST fused to this first EVI-1 zinc finger domain, determined GA(C/T)AAGA(T/C)AAGATAA as a consensus binding sequence (Delwel et al., 1993). The second C-terminal distal domain of three zinc fingers bind a consensus sequence of GAAGATGAG (Funabiki et al., 1994). The ability to

bind GATA-like motifs may be part of the mechanism of inhibition of GATA-1 target genes (Kreider et al., 1993).

Direct evidence of in vivo DNA interaction by ChIP-PCR has been demonstrated at a *Gata-2* upstream enhancer which was described above (Yuasa et al., 2005). To discover further gene targets another group transfected murine fibroblasts with EVI-1 coupled to VP16, a potent trans-activator (Yatsula et al., 2005). Candidate genes were identified by their upregulation on a gene expression array and these included *Gata2*, *Gata3* and *FOG2*. EVI-1 binding near *Gata2* and *FOG2* were confirmed by ChIP-PCR. In an ovarian carcinoma cell line FLAG tagged EVI-1 was immunoprecipitated and subjected to high throughput sequencing (Bard-Chapeau et al., 2012). Although this is in another tissue type, this provides the only reliable genome wide evidence of direct EVI-1 DNA binding capacity to date. 12618 peaks were identified following appropriate validation. EVI-1 bound genes correlated significantly by gene site enrichment analysis to previously published gene expression profiles for *Evi-1* overexpression AML samples and to *Evi-1* KO haematopoietic cells. Binding motifs for GATA (GATAGA) (3178 peaks) and ETS (TCCT/G)(5097 peaks) were identified by de novo motif search, purportedly associated with the two separate DNA binding domains, however, only a minority (6-8%) of the ChIP seq peaks had the two peaks co-occurring. AP-1 motifs were also represented in EVI-1 binding sites with their Co-IP studies suggesting a direct physical interaction between FOS and EVI-1. Significant and previously identified targets include *Gata-2*, *Fos* and *Jun* (Tanaka et al., 1994).

A second EVI-1 ChIP seq experiment in murine myeloid cell line (Glass et al., 2013) was less informative, due to the quality of its ChIP-seq but confirmed *Gata-*2 as a putative target and suggested *Cebpe* might also be a direct EVI-1 target.

EVI-1 associates predominately with transcriptional co-repressors. Reporter gene assays suggests a functional role for the association of EVI-1 with both class 1 and 2 HDAC. One report suggested an association of HDAC1 and 4

(Chakraborty et al., 2001), whilst a second report corroborated the association with HDAC1 (Vinatzer et al., 2001). This may be via a scaffolding protein C-terminal binding protein (CtBP) (Palmer et al., 2001, Shi et al., 2003), which has been shown to be functionally important in the inhibition of the TGFβ signalling pathway (Izutsu et al., 2001).

Histone 3 Lysine 9 trimethylation (H3K9me3) is a well-characterised epigenetic mark of heterochromatin which is important for gene repression and maintenance of nuclear integrity (Shilatifard, 2006). EVI-1 possesses in vitro histone methyl transferase activity (Pinheiro et al., 2012) and has a redundant role in the formation of H3K9me1 in the cytoplasm of cell. Double knockdown of both PRDM3 (*Evi-1*) and PRDM16 led to a loss of H3K9me3 and subsequent breakdown of nuclear lamina. In this light, two independent reports confirming the association of EVI-1 with two histone methyltransferase that play an important role in the formation of H3K9me3: SuV39H1 and G9a is significant (Spensberger and Delwel, 2008, Cattaneo and Nucifora, 2008). A third study suggests that this association is important in the immortalisation of murine bone marrow progenitor cells by *Evi-1* transfection (Goyama et al., 2010).

EVI-1 has the ability to modulate the activity of other transcription factors without directly binding DNA itself. For example, EVI-1 can bind to SMAD3 which normally acts in complex with SMAD4 as part of the TGFβ signalling pathway. EVI-1 bound SMAD3/4 has a reduced DNA binding activity, and thus suppresses TGFβ mediated gene repression. For this activity to occur, the presence of EVI-1 binding sites in the DNA are not required although the EVI-1 zinc finger domains are required (Kurokawa et al., 1998b). Similarly, EVI-1 binds to a variety of other factors: GATA1, reducing its DNA binding capacity (Laricchia-Robbio et al., 2006); the C-terminus of PU.1 preventing appropriate JUN interaction with this transcription factor (Laricchia-Robbio et al., 2009); RUNX1, thereby reducing its DNA binding capacity (Senyuk et al., 2007) and C/EBPα, preventing its auto-regulatory transcriptional activity (Tokita et al., 2007). None of these mechanisms necessarily require direct DNA binding by EVI-1 itself.

In summary, direct evidence of EVI-1 repression at gene targets are scarce despite evidence for EVI-1 as a DNA binding factor and its association with co-repressors. An exception to this is the repression of PTEN by the recruitment of polycomb proteins from the PRC 2/3/4 class to an EVI-1 binding site in the PTEN promoter region (Yoshimi et al., 2011). An association with an increase in Histone 3 Lysine 27 trimethylation (H3K27me3), another well characterised important mark of gene repression important in development, was seen at this genomic locus.

1.6 RUNX1 fusion proteins in myeloid malignancies

Mutations involving *RUNX1* are one of the most commonly found abnormalities in Acute Myeloid Leukaemia (AML). The function of RUNX1 can be disrupted in a number of ways: point mutations (Osato et al., 1999) disrupting the RUNT DNA binding domain, translocations affecting CBFβ which is required for optimal RUNX1 DNA binding (Liu et al., 1993) and translocations of *RUNX1* to other genes resulting in dysfunctional fusion products (Golub et al., 1995, Gamou et al., 1998). The most common fusion protein is RUNX1-ETO, (a product of the t(8;21) translocation) (Miyoshi et al., 1993, Erickson et al., 1992).

The product of another core-binding-factor translocation, t(3;21)(q26;q22), is RUNX1-EVI1. The fusion partner, *EVI-1* was first identified as a target of retroviral insertion in the AKXD mouse strain who develop myeloid leukaemia (Mucenski et al., 1988). *EVI-1 o*verexpression is a common finding in patients with acute myeloid leukaemia (AML) and is an independent poor prognostic factor on multivariate analysis (Groschel et al., 2010). *Evi-1* overexpression in murine bone marrow cells by retroviral infection results in a myelodysplastic condition that transforms into AML (Buonamici et al., 2004).

These next sections will compare these two fusion proteins to elucidate common and distinct properties of core-binding factor fusion proteins.

1.7 Comparison of RUNX1-ETO and RUNX1-EVI-1

In patients with t(8;21) translocations the breakpoints in chromosome 21 clustered to three breakpoint cluster regions in intron 5 (Zhang et al., 2002). The breakpoint in chromosome 8 is clustered around intron 1b resulting in a fusion of the Runt domain of RUNX1 to four conserved Nervy Homology Regions (NHR) domains of ETO (Figure 1-4). The NHR domains are capable of recruiting specific interacting partners and enable RUNX1-ETO to oligomerise with each other (see section 1.7.2). Although there are two zinc fingers in the distal C-terminal end of the fusion protein, to date there is no evidence for DNA binding capability by this domain.

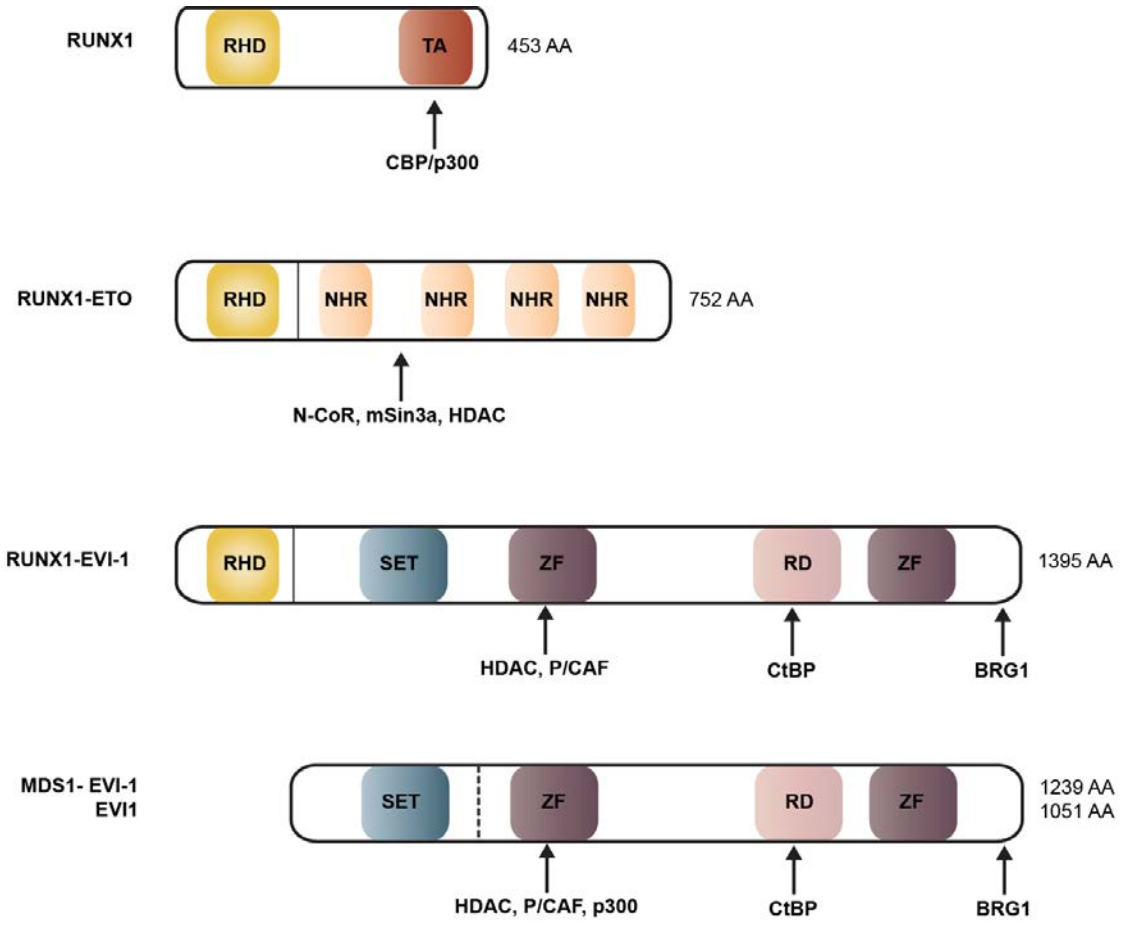


Figure 1-4 Structure of RUNX1, EVI-1, RUNX1-ETO and RUNX1-EVI-1 with their interacting partners
Key AA- amino acids, RHD- Runt homology domain, TA- transactivation domain, NHR- nervy homology region, SET- Su(var)3-9 and 'Enhancer of zeste' and Trithorax, ZF- zinc finger domain, RD- proline rich repressive domain, CBP- CREB binding protein, HDAC- histone deacetylase, CtBP- C-terminal binding protein, N-CoR- nuclear receptor co-repressor.

The t(3;21) translocation results in a similar fusion protein: the RUNT domain is retained in a fusion with the entire *MDS1-EVI-1* gene (Mitani et al., 1994, Nucifora et al., 1994). This includes the SET domain containing PR region in the *MDS* gene. The Su(var)3-9 and 'Enhancer of zeste' and Trithorax(SET) is an evolutionary conserved domain, first identified in drosophila, and has been shown to have histone methyltransferase activity (Pinheiro et al., 2012). Both sets of zinc fingers with direct DNA binding activity (Morishita et al., 1995) are also retained. However, variations in the location of the breakpoint alter the

fusion partner with some examples exchanging *EVI-1* with the upstream gene *EAP*, or the exclusion of the SET domain containing *MDS* gene (Nucifora et al., 1994, Lugthart et al., 2010). Despite the presence of DNA binding domains in the fusion protein, only in vitro direct DNA binding is available (discussed further below).

1.7.1 Clinical description of patients with core-binding factor fusion proteins

Mutations involving *RUNX1* remain the most frequent mutations in patients with AML. Differing point mutations and cytogenetic abnormalities in patients are independent prognostic factors, based on multivariate analyses (Grimwade et al., 2010, Gaidzik et al., 2011). Here, we describe the clinical characteristics of patients with AML expressing the fusion proteins RUNX1-ETO and RUNX1-EVI-1. The distinct clinical characteristics of patients with each type of CBF AML suggest an intrinsic biological difference between the two fusion proteins.

Epidemiology and prognosis

t(8;21) (involving the heterodimeric partner of RUNX1, CBFβ) involves 7% of newly diagnosed younger patients with AML (Grimwade et al., 2010). In patients above 60 years of age, the frequency of these translocations is reduced to 3% (Burnett et al., 2012). The presence of a translocation in AML involving core-binding factor (CBF) protein is universally recognised as a better prognostic factor by a number of independent cooperative groups (Grimwade et al., 2010, Byrd et al., 2002, Slovak et al., 2000). Younger patients with t(8;21) or inv(16) (affecting CBFβ) have a 5 year overall survival of over 84% with advances attributed, in part, to the use of high dose cytarabine in remission consolidation therapy and the use of Gemtuzumab Ozogamacin (humanised anti-CD33 conjugated to calicheamicin, a DNA damaging agent) in remission induction treatment (Hills et al., 2013, Loke et al., 2015).

In contrast, the t(3;21)(q26;q22) which results in the fusion protein RUNX1-EVI-1 is rarely found in patients with de novo AML, with only 9 patients in a cohort of

6515 (Lugthart et al., 2010). This fusion protein is more commonly found in those with therapy related MDS/AML (4 patients in a series of 144) (Rubin et al., 1990) or as a secondary event in the transformation of CML from chronic phase to blast crisis (Rubin et al., 1987, Paquette et al., 2011). The outcomes of patients with secondary AML or CML in blast crisis are poor due to chemo-resistance of the underlying condition. The-5 year event free survival for patients with AML and the presence of RUNX1-EVI-1 fusion is only 14% (Lugthart et al., 2010).

Cooperating mutations

The presence of recurrent cooperating mutations suggests possible synergistic oncogenic mechanisms. Early hypotheses suggested, for the development of AML, a need for two independent mutations: “class I” mutations affecting cell growth and proliferation whilst a “class II” mutation disrupted cellular differentiation and apoptosis (Speck and Gilliland, 2002, Gilliland and Griffin, 2002). In keeping with this original paradigm, recurring additional molecular mutations in t(8;21) patients most frequently cause aberrant activation of cell signalling pathways. The most common cooperating mutation is in the receptor tyrosine kinase KIT, found in 17% of AML patients with the t(8;21) translocation (Pollard et al., 2010, Krauth et al., 2014). *KIT* is highly expressed in haematopoietic stem cells and is normally stimulated by the ligand SCF and is often over expressed in patients with RUNX1-ETO (Paschka and Dohner, 2013). The most common locations are in exon 17 encoding an activation loop, affecting the tyrosine kinase domain and exon 8 encoding an extracellular part of the receptor, affecting receptor dimerisation. Retroviral transduction of *RUNX1-ETO* and either mutations in exon 8 or 17 of *KIT* into murine progenitor cells lead to an AML when transplanted into lethally irradiated mice (Nick et al., 2012). In general, *KIT* exon 17 mutations are associated with decreased event free survival and overall survival on multivariate analyses (Kim et al., 2013, Park et al., 2011). However, it is notable some large studies had also found no significant influence on survival outcomes based on *KIT* mutation subgroup analyses (Pollard et al., 2010). As such *KIT* mutation analysis has not been

incorporated as part of routine clinical practice. However, the German AML SG cooperative group are investigating targeting KIT by the use of tyrosine kinase inhibitor dasatinib, in patients with newly diagnosed CBF AML in a phase III randomised controlled trial (clinical trials identifier NCT02013648).

Other recurring molecular lesions often include other receptor tyrosine kinase mutations, such as in *NRAS, KRAS, FLT3* (both internal tandem duplications and tyrosine kinase domain mutations), *JAK2* V617F and *CBL*. Epigenetic modifiers affected frequently include *ASXL2* (part of the polycomb repressive complex mechanism, discussed above), as well as *IDH1* and 2 (linked to DNA hypermethylation through TET2 disruption) (Krauth et al., 2014). In targeted sequencing approach of 215 CBF leukaemia, 18% of t(8;21) AML had a mutation in the cohesin complex and these patients were at an increased risk of disease relapse after chemotherapy treatment (Duployez et al., 2016). A recently discovered set of mutations cooperating with RUNX1-ETO are missense or truncation mutations in *ZBTB7A* (also known as LRF or Pokemon) (Hartmann et al., 2016). 23% of t(8;21) AML also had a mutation of *ZBTBZ7A*. ZBTB7A is a transcription factor and these mutations appear to disrupt the distal C-terminal zinc finger domains. It has an important role in lymphopoiesis and erythropoiesis as well as the regulation of glycolysis (Lunardi et al., 2013, Liu et al., 2014, Masuda et al., 2016).

Cooperating mutations with RUNX1-EVI-1 are more difficult to identify due to the rarity of the mutation. However, there is a general association of *NRAS* mutations with inv(3) and other 3q26 chromosomal abnormalities, which includes patients with *RUNX1-EVI-1* (Lugthart et al., 2010). In this series of 9 patients with RUNX1-EVI-1: *NPM1*, *FLT3*-ITD and *C/EBPα* mutation were not detected (H Dohner, personal correspondence). In terms of chromosomal abnormalities, out of the 9 patients 4 had monosomy 7, 2 deletion 5q and 2 complex karyotype. The association with development of blast crisis in CML would suggest a corroborative function with the BCR-ABL fusion protein. This is supported by murine studies which show that both *BCR-ABL* and *RUNX1-*

EVI-1 transduction in progenitor cells are required for the development of an AML phenotype (Cuenco and Ren, 2001, Cuenco and Ren, 2004).

1.7.2 RUNX1-ETO role in development of leukaemia

Additional cooperative mutations are required for the development of leukaemia

A number of independent lines of enquiry suggest that RUNX1-ETO expression alone is not sufficient to cause AML and require the acquisition of additional driver mutations. Firstly, it is possible to isolate *RUNX1-ETO* transcripts from patients in long term remission (Miyamoto et al., 2000). Stringent FACS sorting of haematopoietic cells demonstrate detectable levels of *RUNX1-ETO* transcripts in HSC, monocytes and B cells but not in T cells. Secondly, Guthrie spot analyses allow the identification of mutations that had occurred during foetal development (Wiemels et al., 2002). Despite the detection of the *RUNX1-ETO* translocation in neonatal blood, the development of AML occurred in patients with a delay of up to 10 years, and in some patients these transcripts were detectable long into remission. Thirdly, transgenic mice with *RUNX1-ETO* driven from a myeloid promoter did not develop AML during their lifetime. However, they were predisposed to the development of AML when they were exposed to an alkylating agent, *N*-ethyl-*N*-nitrosurea compared to their wild-type littermates (Yuan et al., 2001).

RUNX1-ETO is a dominant negative inhibitor of RUNX1

Two murine models suggest that RUNX1-ETO behaves as a dominant negative inhibitor of RUNX1 (Yergeau et al., 1997, Okuda et al., 1998). Knock in of *RUNX-ETO* in the *Runx1* locus results in a constitutive heterozygous *RUNX1-ETO* /*Runx*1 genotype. These mice phenotypically resemble mice with *Runx1* knockout: they die at E13.5 with lack of normal fetal liver haematopoiesis and haemorrhages in the central nervous system, suggesting RUNX1-ETO disrupted native RUNX1 function. This is supported by the use of gene reporter studies that utilise the GM-CSF promoter. The ability of native RUNX1 to trans

activate a CAT reporter assay is abrogated by the presence of RUNX1-ETO (Frank et al., 1995). RUNX1-ETO can also directly bind RUNX motif containing sequences in *in vitro* assays (Meyers et al., 1993) Genome-wide studies involving the depletion of RUNX1-ETO by siRNA has been accomplished (Ptasinska et al., 2012). After siRNA knock down RUNX1-ETO and subsequent RUNX1 ChIP seq, it can be confirmed a third of binding sites occupied by RUNX1 were previously occupied by RUNX1-ETO.

The importance of wildtype RUNX1 in maintaining survival of cells with RUNX1-ETO has been seen in two studies (Ben-Ami et al., 2013). One study showed that knockdown of RUNX1 induced apoptosis in t(8;21) Kasumi-1 cell line, that was rescued by knockdown of RUNX1-ETO. This is in part due to the previously recognised finding that RUNX1-ETO is pro-apoptotic (Burel et al., 2001), and the wild-type RUNX1 counteracts this. Furthermore, RUNX1 appears to directly regulate a set of mitotic control genes that enables continued proliferation of these leukaemic cells (Ben-Ami et al., 2013). In a second study, the authors suggest the presence of RUNX1 is a general pro-survival mechanism in other forms of AML, mediated through BCL2 (Goyama et al., 2013).

RUNX1-ETO increases proliferation and self-renewal properties of haematopoietic cells

Due to the embryonic lethality of constitutive *RUNX1-ETO* knock-in transgenic mice, a conditional *RUNX1-ETO* knock in transgenic mouse provides a model for the direct role of the *RUNX1-ETO* translocation (Higuchi et al., 2002). Cre mediated recombination resulted in the activation of the *RUNX1-ETO* allele through the deletion of a loxP flanked transcriptional stop cassette. The result is a *RUNX1-ETO* allele driven by the endogenous *Runx1* promoter, achieving physiologically expected mRNA levels in a lineage specific manner. This model demonstrated that in long-term replating assays, that there was increased proliferation and a capacity for self-renewal. However, there was no striking impediment to the ability to differentiate; with the visible presence of mixed lineage colonies. This is supported by reports described above, where, despite

the presence of RUNX1-ETO in HSCs multi-lineage haematopoiesis remains. Ectopic expression of RUNX1-ETO in human CD34+ stem and progenitor cells isolated from cord blood results in their increased self-renewal properties (Mulloy et al., 2002).

A recent, more refined model, with conditional expression of RUNX1-ETO in early haematopoietic progenitors have shown that the fusion protein can expand Granulocyte-Macrophage Progenitors (GMP), whilst HSC numbers are largely unaffected (Cabezas-Wallscheid, 2013). These RUNX1-ETO expressing GMP cells have leukaemia initiating potential when transplanted into lethally irradiated mice. The long latency in developing AML in this model once again emphasised the need for additional mutations in combination with RUNX1-ETO, in order to initiate leukaemia.

RUNX1-ETO inhibits myeloid differentiation in fully transformed leukaemic cells

In a U937T cell line expressing tetracycline inducible *RUNX1-ETO,* induction of the fusion protein leads to a pro-apoptotic stage with reduced neutrophilic differentiation ability (Burel et al., 2001). Furthermore, siRNA knockdown of RUNX1-ETO results in differentiation of the leukaemic blasts into more mature myeloid cells (Dunne et al., 2006, Ptasinska et al., 2012) and in a doxycycline inducible model of RUNX1-ETO, withdrawal of doxycycline and thus downregulation of RUNX1-ETO led to a decrease in leukaemic cell number (Cabezas-Wallscheid, 2013). These experiments suggest that the fusion protein is necessary for fully transformed cells to retain their leukaemic phenotype.

Explanations for the phenotypic variations seen in these different models suggest several possibilities. Firstly, dosages of the fusion protein appears to be a major difference; if the fusion protein is driven by the native *Runx1* promoter, the doses of the fusion protein would be lower, particularly in later, more differentiated lineages. Secondly, the cellular context and content of the

transcription factor pool would be important and thus which progenitor type (e.g. HSC, GMP etc.) RUNX1-ETO is expressed in, would be crucial. Finally, the presence of co-existing mutations may be required to reveal the clear differentiating block seen in t(8;21) AML.

How might this differentiation block be affected? C/EBPα is a key transcription factor of the myeloid lineage, directly regulating many genes involved in the differentiation of cells into mature granulocytes; which are absent in C/EBPα mutant mice (Zhang et al., 1997b). C/EBPα also has a role in maintaining HSC quiescence (Ye et al., 2013). Investigations have demonstrated RUNX1-ETO down regulates C/EBPα through two ways: the fusion protein can physically bind to C/EBPα, thereby preventing its ability to activate transcription of C/EBPα target genetic elements, as demonstrated through a luciferase reporter assay system (Westendorf et al., 1998). This is important because C/EBPα activates its own promoter through an upstream stimulatory factor binding site (Timchenko et al., 1995, Legraverend et al., 1993). It has been shown that expression of *CEBPA* itself is reduced through the inhibition of this autoregulatory mechanism, by RUNX1-ETO interaction with C/EBPα itself and not directly at the promoter of *CEBPA*. This results in a reduction in C/EBPα levels, demonstrable in both cell lines and patients with the t(8;21) translocation (Pabst et al., 2001). It has been shown that enforced *C/EBPα* expression in the t(8;21) Kasumi-1 cell line can rescue neutrophilic differentiation (Pabst et al., 2001). Finally, RUNX1-ETO may also regulate *CEBPA* through binding to its upstream enhancer (Avellino et al., 2016, Ptasinska et al., 2014).

Recruitment of co-repressors by RUNX1-ETO represses target genes

The loss of the transactivation domain of the native RUNX1 and its replacement by ETO is associated with de-acetylation of histone 3 lysine 9 on target genes (Ptasinska et al., 2012, Martens et al., 2012), which should result in reduced chromatin accessibility and repression of gene expression (Bannister and Kouzarides, 2011). This probably occurs via co-repressor complex associated

with RUNX1-ETO such as N-CoR/mSin3/SMRT/HDAC. Interaction of RUNX1-ETO with N-CoR was initially identified through the use of a yeast-two-hybrid library screen (Wang et al., 1998) and the recruitment of SMRT to RUNX1-ETO subsequently shown to be directly dependent on the NHR domains (Liu et al., 2007). The ability of RUNX1-ETO to repress transcription has been shown to be dependent on its ability to associate with N-CoR (Lutterbach et al., 1998). The ability of RUNX1-ETO to recruit HDAC appears to be dependent on the presence of N-CoR (Gelmetti et al., 1998), loss of N-CoR association resulted in an inability to inhibit differentiation in a myeloid cell line model. Depletion of RUNX1-ETO by siRNA leads to a return of histone 3 lysine 9 acetylation (Ptasinska et al., 2012, Martens et al., 2012) on a genome wide basis.

RUNX1-ETO oligomerises and cooperates with other transcription factors in binding DNA

Of the four NHR(1-4) in the ETO fusion partner, NHR 2 plays a crucial role in oligomerisation of RUNX1-ETO. By x-ray crystallography, it was demonstrated that RUNX1-ETO forms a tetrameric structure, which is extremely stable and required extensive mutagenesis to disrupt its structure (Liu et al., 2006). RUNX1-ETO carrying mutant NHR2 domains which were unable to oligomerise, had compromised self-renewal capacity and ability to inhibit gene expression. Although, mutations that inhibited oligomerisation had no direct impact on the affinity of RUNX1-ETO to other co-repressors, the four-fold increase in surface area by the oligomer is likely to be important in the recruitment of other proteins (Liu et al., 2006, Sun et al., 2013).

Other than co-repressors, RUNX1-ETO acts cooperatively with other direct DNA binding transcription factors. Examples of these include ETS family proteins, which also interact with the wild-type RUNX1, and this supported by the co-localisation of these binding motifs with RUNX1-ETO ChIP seq (Ptasinska et al., 2012, Martens et al., 2012). Specific examples of this family include activating transcription factor ERG and FLI1. ChIP-seq data for ERG, FLI-1 and RUNX1-ETO show strong overlap in their binding sites. Through sequential re-ChIP experiments with first ERG and then RUNX1-ETO and also,

vice versa, it can be shown that they bind together at the same loci. In a model where RUNX1-ETO is inducible, it was shown that FLI-1 bind to genomic loci prior to the binding of RUNX1-ETO (Martens et al., 2012).

Another group of transcription factors which are bound by RUNX1-ETO are the E-protein family. E proteins are widely expressed, and play important roles in cellular development (Lassar et al., 1991). Mass spectrometric analyses of RUNX1-ETO bound complexes identified the E protein family member HEB (Zhang et al., 2004a). In RUNX1-ETO the NHR1 domain bears extensive homology to TBP associated factor protein which is part of the TFIID transcription factor complex. This domain was shown to be crucial in maintaining a high affinity interaction with HEB, and displaces co-activating histone acetylase complex p300/CAF from their binding site on the Activation Domain 1 (AD1) on HEB. These results in suppression of HEB activated promoter activity. The related E protein E2A is also an interacting partner of RUNX1-ETO but is bound with a lower affinity. On a genome wide basis motif analyses show an enrichment of E-Box sites at RUNX1-ETO ChIP seq peaks (Ptasinska et al., 2012) (Sun et al., 2013). It was later shown the oligomerised RUNX1-ETO may be important for the interaction with E proteins (Sun et al., 2013) but this is not consistent with earlier data described above (Liu et al., 2006).

In addition, further analyses of RUNX1-ETO complexes identified LMO2 (with its partner LDB1) and LYL1. ChIP seq of RUNX1-ETO, HEB, E2A and LMO2 show shared genomic loci. ShRNA knockdown of these components (i.e. HEB/E2A, LY1, LDB1 and LMO2) reduced the lethality of RUNX1-ETO induced leukaemia in mice (Sun et al., 2013).

1.7.3 Comparison of the role of RUNX1-EVI-1 and RUNX1-ETO in the pathogenesis of leukaemia

The two fusion proteins contain different C-terminal fusion partners and have different functional activities. Notably, *ETO* is not a commonly found oncogenic

mutation in its own right, whilst as described already; *EVI-1* is overexpressed in AML. This is reflected in murine models where overexpression of *Evi-1* results in a MDS phenotype (Buonamici et al., 2004). In contrast, *Evi-1* knockout mice demonstrate reduced leukaemic potential when transfected with potent oncogenic fusion proteins such as E2A-HLF and MLL-ENL (Goyama et al., 2008). However, the oncogenic potential of the fusion protein RUNX1-EVI-1 is not only a result of a loss of RUNX1 function and gain of EVI-1: overexpression of EVI-1 in RUNX1 -/- murine bone marrow cells fails to transform the murine bone marrow cells, whilst the full length fusion protein was able to (Takeshita et al., 2008).

Similarities between RUNX1-EVI-1 and RUNX1-ETO

RUNX1-EVI-1 promotes an acute myeloid leukaemia but requires other mutations

When mice were transplanted with bone marrow cells transduced with *RUNX1-EVI-1*, they developed leukaemia within 5-13 months. There was a reduced latency with subsequent serial transplants. This suggests other cooperative mutations are required to accumulate with time before the onset of a full leukaemia (Cuenco et al., 2000).

RUNX1-EVI-1 is found in secondary AML and CML in blast crisis (Rubin et al., 1990, Rubin et al., 1987, Paquette et al., 2011) and is therefore likely to be a secondary driver mutation and not an initiating mutation. Recent single cell gene expression analysis has confirmed that *RUNX1-EVI-1* is a secondary mutation in a patient with CML in blast crisis (Nukina et al., 2014).

RUNX1-EVI-1 acts as a dominant negative inhibitor of native RUNX1

RUNX1-EVI-1 knocked into the *RUNX1* locus initiates a similar phenotype to the RUNX1-ETO knock in mice (Maki et al., 2005) who die at E13.5 with a failure of adult haematopoiesis. Haematopoietic cells from the fetal liver differentiate into dysplastic megakaryocytes and myeloid progenitors.

Like RUNX1-ETO, RUNX1-EVI-1 can also bind RUNX motif containing sequences, and does so with a higher affinity than native RUNX1 (Tanaka et al., 1995). This may be due to increased binding affinity to CBFβ: in a co-transfection model CBFβ and RUNX1-EVI-1 co-localise in the nucleus, by Co-IP and fluorescent microscopy (Tanaka et al., 1998).

Finally, RUNX1-EVI-1 may bind to RUNX1 itself in Co-IP experiments using HA and FLAG tagged constructs (Senyuk et al., 2007). This was confirmed by immunofluorescent microscopy, when 293T cells were transiently transfected by RUNX1-EVI-1 and RUNX1. This ability is mediated through a distal EVI-1 portion of the fusion protein as binding to RUNX1 is also a property that is conserved with EVI-1 and the isolated distal fragment.

RUNX1-EVI-1 promotes malignant proliferation and self-renewal of affected cells

RUNX1-EVI-1 knock in transgenic murine model (Maki et al., 2005), was embryonically lethal, with severely impaired fetal liver haematopoiesis. The few haematopoietic progenitors that could be isolated from the fetal liver were dysplastic in appearances and demonstrate a significant increase in in vitro replating capacity.

RUNX1-EVI-1 may inhibit differentiation through down-regulation of C/EBPα activity

Fetal liver haematopoietic cells from *Runx1-Evi-1* knock in transgenic murine model described above (Maki et al., 2005) had impaired differentiation, with a bias towards macrophage development in colony plating assays. Murine myeloid progenitor LG-3 cells differentiate into granulocytic cells when they are exposed to G-CSF. RUNX1-EVI1 is capable of inhibiting differentiation in this model (Tokita et al., 2007) but differentiation could be rescued by C/EBPα co-

expression. Direct binding of RUNX1-EVI1 to C/EBPα could be observed and this decreased C/EBPα DNA binding ability in EMSA and in luciferase reporter assays (Tokita et al., 2007). It is unclear whether this results in a reduction in *C/EBPα* mRNA levels through inhibition of *C/EBP*α autoregulatory mechanism.

RUNX1-EVI-1 Inhibition of C/EBPα may also occur via calreticulin. Calreticulin is a multifunctional protein that can act as a chaperone for misfolded proteins but is also involved in intracellular signalling through sequestration of calcium ions in the endoplasmic reticulum (Luo and Lee, 2013). In U937 cells transfected with RUNX1-EVI-1 and human patient samples, C/EBPα protein levels were disproportionately lower than the mRNA levels would suggest. Calreticulin was more highly expressed than expected in patients with RUNX1-EVI-1 and this was recapitulated when RUNX1-EVI-1 was transfected in to a cell line. Calreticulin may bind to *C/EBPα* mRNA, and prevent its translation (Timchenko et al., 2002). Knockdown of calreticulin by siRNA restored C/EBPα levels (Helbling et al., 2004).

RUNX1-EVI-1 oligomerisation

Oligomerisation of oncoproteins improves their ability to recruit other active proteins. Like RUNX1-ETO, RUNX1-EVI-1 can oligomerise with other molecules of the fusion protein (Senyuk et al., 2005). A series of deletion mutants affecting different portions of the fusion protein showed that a series of domains, ranging from the RUNX1 through to both the proximal and distal Zinc finger portions are required for this self-interacting property. Although deletion mutants of the distal zinc finger portions of RUNX1-EVI-1 impaired the ability of this fusion protein to inhibit differentiation (Senyuk et al., 2005), it is unclear whether this is specifically related to the loss of oligomerisation property of the fusion protein or whether this is due to the loss of other properties mediated by this portion of the fusion protein.

Differences between RUNX1-EVI-1 and RUNX1-ETO

Although both CBF fusion proteins require additional mutations before the onset of an overt leukaemia, RUNX1-EVI-1 promotes a more aggressive leukaemia than RUNX1-ETO. Mice transplanted with bone marrow cells expressing *RUNX1-EVI-1* have a reduced latency to leukaemia (Cuenco et al., 2000) compared to bone marrow cells expressing *RUNX1-ETO* (Schessl et al., 2005, Schwieger et al., 2002). The onset of leukaemia in *RUNX1-EVI-1* transgenic mice (Maki et al., 2006) do not require administration of alkylating agents, unlike *RUNX1-ETO* transgenic mice (Yuan et al., 2001), or in comparison at least have reduced latency to development of leukaemia (Cabezas-Wallscheid, 2013).

Some differences between the two CBF fusion proteins are likely to be attributable to the C-terminal fusion partner EVI-1. As described above, these include stimulation of the AP-1 pathway (Tanaka et al., 1995) and inhibition of the TGFβ pathway (Kurokawa et al., 1998a, Sood et al., 1999) through the Smad3 protein. With the native EVI-1 this is also mediated via the repressive complex CtBP containing HDAC (Izutsu et al., 2002).

RUNX1-EVI-1 binds to CtBP as a co-repressor

CtBP is a multifunctional co-repressor that can act as a scaffolding protein. It exists in a complex that contains histone deacetylases (HDAC1), polycomb repressive complexes (HPC2), and H3K9 methyltransferases (G9a) (reviewed in (Chinnadurai, 2007)). All of these enzymes promote a condensed chromatin configuration. For example, in a cell line, in which the E-cadherin gene is not expressed, CtBP1 localises at the e-cadherin promoter and recruits HDAC1 and 2, HPC2, and G9a. This resulted in methylation of H3K9 and loss of H3K9 acetylation (Shi et al., 2003).

RUNX1-EVI-1 has been reported to recruit CtBP which was suggested to be in conjunction with inhibition of *C/EBPα* transcription (Tokita et al., 2007) and

RUNX1 mediated transactivation (Izutsu et al., 2002). However, these experiments have only been performed through reporter assays, and the direct epigenetic mechanisms have not been elucidated.

1.8 Aims of the project

Although both CBF fusion proteins result from a translocation of the *RUNX1* gene it is unclear why the two leukaemia have such strikingly different clinical outcomes. Understanding the precise mechanism of how the two CBF fusion proteins programs the epigenome will likely result in an explanation for this difference in prognoses of patients with these two mutations. This thesis therefore addresses the following open questions:

How does the gene regulatory cistrome compare between RUNX1-ETO and RUNX1-EVI-1 expressing leukaemias?

The two types of CBF AML differ greatly in their clinical characteristics. It is unknown how this is reflected by differences in gene expression and the associated regulatory cistrome. In this project we will characterise the epigenome, using RNA-seq and DNase-seq, of primary t(3;21), t(8;21) AML cells, as well as normal PBSCs.

How is RUNX1-EVI-1 directed to gene targets?

RUNX1-ETO direct DNA binding ability has been demonstrated in vivo in patients and original cell lines from patients with this translocation. DNA binding activity for RUNX1-EVI-1 has only been demonstrated by in-vitro studies and transfected cell lines in which the fusion protein has been over expressed. We will identify RUNX1-EVI-1 binding targets by ChIP-seq and compare RUNX1-EVI-1 binding sites to RUNX1-ETO targets.

What is the role of wild-type RUNX1 in t(3;21)AML and how does this compare to its role in t(8;21) AML?

Wild-type RUNX1 plays a vital role in RUNX1-ETO driven AML and the biological implication and mechanism for this has been carefully elucidated. In t(3;21) leukaemia the binding pattern of RUNX1 as compared to RUNX1-EVI-1 is unknown. We aim to characterise RUNX1 binding sites in t(3;21) leukaemia by ChIP seq and compare these sites to RUNX1-EVI-1 targets in t(3;21) leukaemia as well as to RUNX1 binding sites in t(8;21) cells and in normal CD34 PBSCs.

Is RUNX1-EVI-1 required to maintain the leukaemic phenotype of t(3;21) cells?

RUNX1-ETO is required to maintain the leukaemic phenotype of t(8;21) AML cells. However, the t(3;21) translocation is often a secondary event in the transformation of leukaemic cells and it is unknown whether the expression of RUNX1-EVI-1 is required to maintain its leukaemic phenotype. We will characterise the changes to t(3;21) cells after RUNX1-EVI-1 knockdown by siRNA.

What are the shared epigenetic mechanisms between the two CBF fusion proteins?

Our group has previously documented the genome wide epigenetic reprogramming events that follow knock down of RUNX1-ETO in t(8;21) cells (Ptasinska et al., 2012, Ptasinska et al., 2014). In this project we will identify the epigenetic re-programming events following RUNX1-EVI-1 knockdown in t(3;21) cells and identify common features seen following the knockdown of both CBF fusion proteins that may enable targeting of both CBF leukaemias.

Chapter 2. Methods and Materials

2.1 Cell line culture

Cells were maintained in a humidified incubator at 37°C with 5% CO_2. t(3;21) SKH-1 and K562 cells were cultured in RPMI medium with 10% fetal calf serum (FCS) supplemented with glutamine and penicillin/streptomycin. t(8;21)Kasumi-1 cells were cultured in RPMI with 15% FCS supplemented with glutamine and penicillin/streptomycin. HEK293T and HeLa cells were cultured in Dulbecco Modified Eagle Media (DMEM) with 10% FCS supplemented with glutamine and penicillin/streptomycin.

2.2 Purification of blood samples from patients with AML

Blood from t(3;21) patient 1 was diluted 1:1 with PBS and layered onto density gradient medium (Lymphoprep, Stem Cell technology, USA). The blood-PBS-Lymphoprep mix was subsequently centrifuged at 592xg (acceleration setting 4, no brakes). After centrifugation, the mixture had separated into two phases with the mononuclear cells separated into a layer in between. This midlayer was then isolated and incubated with 500 µl CD34+ microbeads (Miltenyl-Biotech, USA) for 15 minutes, CD34 microbeads and cells were subsequently suspended in 5ml MACS buffer (Phosphate Buffer Saline (PBS) with 2mM EDTA and 0.5% BSA) and centrifuged at 300xg at room temperature. The cell pellet was isolated and resuspended in 500 µl of MACS buffer and passed through a LS column (Miltenyl-Biotech, USA) placed in a magnetic field, which has been pre-rinsed with 2ml MACS buffer. With the column still on the magnetic field, it was rinsed with thrice with 3ml MACS buffer. The CD34+ cell fraction was isolated by removing the column from the magnetic field and flushed with 5ml of MACS buffer. The CD34+ expression on this fraction was confirmed by flow cytometry before either immediate use for DNaseI hypersensitivity site mapping or RNA extraction by Trizol, as described below.

Cells from t(3;21) patient 2 was previously isolated by density gradient medium and cryopreserved at the Erasmus University Medical Centre, Netherlands. CD34+ cells were thawed with pre-warmed RPMI-1640 +10% FCS. After centrifugation, the cell pellet was re-suspended in 750 µl MACS buffer (PBS with 0.5% BSA and 2mM EDTA) and 35µl CD34-PE, with a separate sample stained with IgG PE as an isotype control. CD34+ cells were isolated by FACS using a MoFlo Astrios (Beckman Coulter, USA) and were then directly used for DNaseI hypersensitivity site mapping or underwent RNA extraction by Trizol, as described below.

2.3 Purification of CD34+ mobilised peripheral blood stem cells

CD34+ peripheral blood stem cells (PBSCs) from healthy adults were mobilised into the peripheral circulation by administrating donors with pegylated G-CSF (trade name: Lenograstim, Chugai Pharmaceuticals, Japan), Suitable patient donors were identified by the NHS Blood and Transplant service (NHSBT). Cells were harvested from the patients by apheresis and stored by NHSBT in liquid nitrogen. Cryopreserved cells were thawed at 37ºC using a waterbath and eluted from storage bag with a PBS/Glucose/Citrate solution (0.09% glucose + 3.3% FCS + 1mM sodium citrate). After centrifugation at 300xg for 5 minutes, the cell pellet was treated with DNaseI (Roche, Switzerland) at 0.6mg/ml concentration in PBS/Glucose/Magnesium/Calcium solution (PBS+0.5 mM $MgCl_2$ + 1.2 mM $CaCl_{2)}$ + 1 % FCS + 0.1 % glucose + 2 mM $MgCl_2$) for 5 minutes at room temperature. Following DNaseI treatment the cells were once again diluted with PBS/Glucose/Citrate solution and the mononuclear layer isolated by density gradient medium and CD34+ beads separation by MACs columns, as described above.

Table 2-1 Details of patient samples.
Details of patient samples included in this study. t(8;21) patient 1 and 2 was processed before the commencement of this study.

Lab name	ID	White cell count	Separation method	Age	Sex	Stage of treatment	Cytogenetics	Clinical notes	Mutations
Birmingham GT027	t(3;21) 1	23	MACS	44	F	Presentation	46,XX,t(3;21)(q26;q22),der(5)t(5;13)(q2;q3),der(7)t(1;7)(q3;q3[5]/47,idem,+12,+der(21)t(3;21)[4]	Therapy related AML (previous Myelofibrosis and T cell lymphoma)	KRAS
Rotterdam 5354	t(3;21) 2	54	FACS	72	M	Presentation	46,XY,t(3;21)(q26;q22),del(12)(p12p13)[20]	RAEB-t	DNMT3, SRSF2
H12812	t(8;21) 1	2.12	MACS	45	M	Presentation	46,XY,t(8;21)(q22;q22)	AML	-
H18909	t(8;21) 2	53	MACS	53	M	Presentation	46,XY,t(8;21)(q22;q22)	AML	CBL, FLT3 TKD

J209/R423	PBSC 1	N/A	MACS	51	M	N/A	N/A	Mobilised PBSCs- sibling donor for allogeneic stem cell transplant	N/A
J299/R454	PBSC 2	N/A	MACS	47	F	N/A	N/A	Mobilised PBSCs- autologous stem cell transplant (CNS lymphoma)	N/A

2.4 Culturing CD34+ mobilised peripheral blood stem cells

Human CD34+ peripheral blood stem cells (PBSC) were maintained with Iscove Modified Dulbecco Media (IMDM) with 20% BIT9500 serum substitute (Stem Cell Technologies, US) with IL-3, IL-6, SCF, TPO, FLT3-L at 10ng/ml and SR1 1μM (Stem Cell Technologies, US).

2.5 siRNA mediated depletion of RUNX1-EVI-1

$1x10^7$ cells were electroporated using an EPI 3500 (Fischer, Germany) electroporation at 350v, 10ms. siRNA sequences (Axolabs, Germany) specific for the translocation breakpoint of Runx1-Evi-1 were 5'-GAACCUCGAAAUAAUGAGUGU-3' (sense) and 5'-ACUCAUUAUUUCGAGGUUCUC-3' (antisense). Control siRNA was 5'-CCUCGAAUUCGUUCUGAGAAG-3' (sense) with 5'-UC UCAGAACGAAUUCGAGGUU-3' (antisense). siRNA was used at 200nM. After electroporation, the cells remained in their cuvettes for 5 minutes before being directly added to RPMI-1640 with 10% FCS, supplemented with penicillin/streptomycin and glutamine at a concentration of 0.5 $x10^6$ cells per ml and returned to an incubator kept at 37°C and 5% CO_2.

2.6 RNA extraction

Pelleted cells from primary patient material were lysed by adding 1ml Trizol ™ (Life Technologies, US). 200μl of chloroform was added and the mixture was manually shaken for 15 seconds. The mixture was incubated at room temperature for 3 minutes. The mixture was centrifuged at 12000xg for 15 minutes at 4°C. The top clear aqueous phase was removed and placed in a fresh tube. 0.5ml of 100% isopropanol was added to the isolated aqueous phase and incubated at room temperature for 10 minutes which was then transferred to a RNeasy Minielute column (Qiagen, USA) and centrifuged for 15s at 8000xg. 350μl of RWI buffer from the RNeasy Kit was added to the column and centrifuged for 15s at 8000xg. DNaseI 10μl and 70μl RDD buffer

(Qiagen, USA) were mixed and added to the column and incubated for 15minutes at room temperature. Afterwards, 350µl of RWI buffer from RNeasy Kit (Qiagen, USA) was added to the column and centrifuged for 15s at 8000xg. Following this 500µl of RPE buffer was added and centrifuged for 15s at 8000xg. The column was washed with 500µl 80% ethanol and centrifuged at 2 minutes at 8000xg. The column was dried by centrifugation at 5 minutes at 8000xg. RNA was eluted from the column by adding 12µl of water to the column followed by centrifugation at 5 minutes at 8000xg.

RNA was isolated from SKH-1 cells by Trizol™ (Life Technologies, US) as by manufacturer's instructions. At the last step of the protocol RNA was resuspended in 17µl of RNAse free water to which was added 2µl of 10x buffer supplied with the Ambion Turbo DNaseI (Thermos Scientific, USA), of which 1 µl was added. All of which was incubated at 37ºC for 30 minutes. The RNA solution was then purified using a Nucleospin RNA clean up column (Machery Nagel, France), according to their instructions. The quality of RNA from all methods was assessed using a spectrophotometer, by the ratio of the absorbance at 260nM and 280nM wavelengths. RNA has a greater absorbance in the 260nM wavelength, Eukaryotic Total RNA PICO Bioanalyser chip (Agilent technologies, USA) allows visualisation of the size of the RNA molecules and thus, demonstrates whether the sample is degraded or not.

2.7 RNA Seq library

RNA-seq libraries were prepared with a Total RNA Ribo-zero library preparation kit (with ribosomal RNA depletion) (Illumina, USA) according to manufacturer's instructions with the following alterations: 15 cycles of PCR was undertaken to amplify the library and adaptors for multiplexing were used at a 1:4 dilution. Library quality was checked by running the samples on a Bioanalyser and libraries were quantified using a Kapa library quantification kit (Kapa Biosystems, USA) and run in a pool of eight indexed libraries in two lane of a HiSeq 2500 (Illumina, USA) using rapid run chemistry with 100bp paired end reads.

2.8 cDNA synthesis

1µg RNA was used to make cDNA with 0.5 µg OligoDT primer, Murine moloney reverse transcriptase and RNase Inhibitor (Promega, USA) according to the manufacturer's protocol.

2.9 Real-time polymerase chain reaction

RT-PCR was performed using Sybr Green mix (Applied Biosystems, UK), at 2x dilution. Primers were used at 100nM final concentration. cDNA was diluted either 1:10 or 1:50 depending on expression levels of targets. A 7900HT system (Applied Biosystems, UK) was used to perform qPCR. Analyses were performed in technical duplicates using a standard curve derived from RNA purified from the untreated cell line (1:10 followed by 1:5 dilutions). Primer sequences are listed in the Appendix.

2.10 Dead cell removal and Annexin V/PI staining for flow cytometry

Dead cell removal was performed using negative selection on a MS column following incubation with Dead Cell Removal microbeads (Miltenyl Biotech, USA) as per manufacturer's instructions. Dead cell removal was performed on all samples prior to RNA extraction or DHSs mapping. Annexin V-APC/PI staining (Ebiosciences, USA) or Annexin V-FITC/PI staining (BD Biosciences, USA) was performed according to manufacturer's instructions. Annexin V-APC staining was used for cells that expressed GFP (e.g. CEBPA-ER retrovirus transduced SKH-1), thereby preventing quantification of Annexin V-FITC surface staining, which were used for all other experiments.

2.11 DNaseI hypersensitivity site mapping and size selection

Prior to DNaseI digestion, apoptotic cells were removed using the Dead Cell Removal Kit (Miltenyl Biotech, UK) as per manufacturer's instructions. 3x 10^7 SKH-1 cells were suspended in 1ml DNase I buffer (0.3M sucrose, 60 mM KCl,

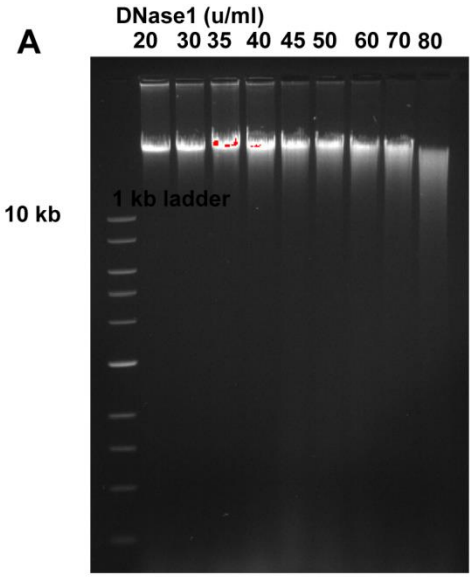
15 mM NaCl, 5 mM MgCl2, 10 mM Tris pH7.4). Digestion on $4.5x10^6$ cells was performed with DNase I (Worthington, DPPF grade) at 80 units/ml in DNase I buffer with 0.4% NP-40 and 2mM CaCl2 at 22ºC for 3 minutes. The reaction was stopped with cell lysis buffer (0.3M NaAcetate, 10mM EDTA pH 7.4, 1% SDS) with 1mg/ml Proteinase K and incubated at 45ºC overnight.

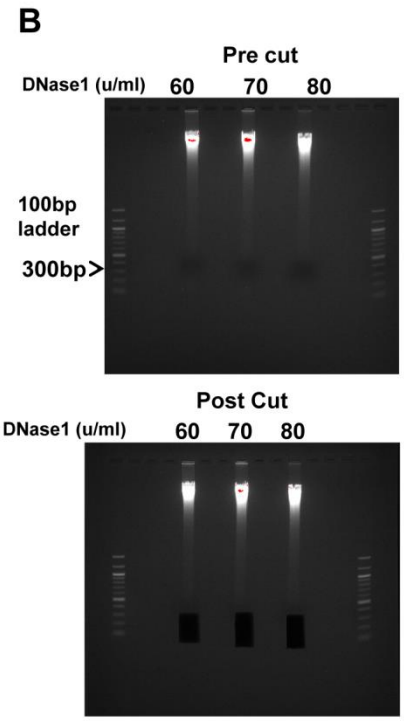
For DHS mapping in CD34+ purified t(3;21) patient cells and in SKH-1 transfected with siRNA, lower cell numbers were available and therefore the DNaseI concentrations were reduced according to the cell numbers available:

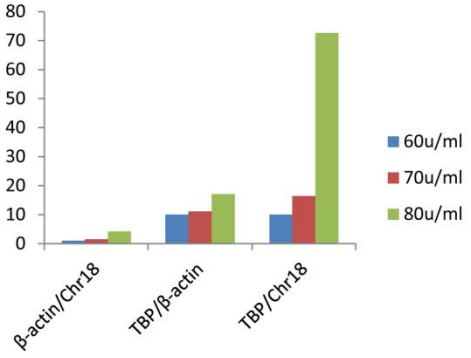
Sample	Cell number per digestion	Final concentration of DNaseI (units/ml)
t(3;21) patient 1	$1.1x10^6$	9
t(3;21) patient 2	$350x10^3$	3
SKH-1 transfected with siRNA	$2.25x10^6$	35 or 45

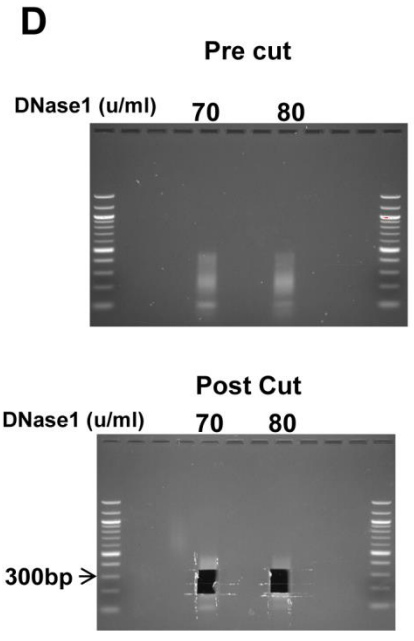
The digested DNaseI material was treated with RNAseA (Sigma Aldrich, Germany) at a final concentration of 100µg/ml at 37ºC for 1 hr. Genomic DNA was extracted using phenol/chloroform method: an equal volume of phenol was added to the reaction and placed on a rotator wheel for 45 minutes. This was centrifuged for 5 minutes at 16000xg at room temperature. The top layer was transferred to a new tube and the process was repeated sequentially with phenol/chloroform and chloroform. After purification by chloroform extraction, genomic DNA was precipitated with ethanol. This was pelleted by centrifugation for 5 minutes, at 16000xg at 4°C. The pellet was resuspended with 70% ethanol and centrifugation for 5 minutes, at 16000xg at 4°C. The pellet was air-dried and dissolved by Tris-EDTA (40mM Tris Acetate 1mM EDTA).

DNaseI preferentially cleaves DNA in open chromatin regions (figure 1-2), but samples can be over digested by the nuclease, whereby heterochromatin and gene body regions may become digested as well. Digestion was checked visually by running the samples on a 0.7% agarose gel (Figure 2-1 A), and by RT-PCR evaluating the ratio of open (TBP promoter) to closed regions of DNA (chromosome 18) and active gene body (beta-actin) to prevent selection of over digested samples (primers in appendix). Optimal digestion is seen in samples where qPCR signal of TBP promoter is high as compared to the β-actin gene body or heterochromatin represented by chromosome 18. Subsequently, between 2 to 10µg of DNaseI digested DNA (depending on material available) were run on a 1.2% agarose gel for selection of shorter fragments to increase the fraction of fragments captured from DHSs. This is because nuclease cleavage is increased in open chromatin regions and therefore releases more short fragments, whilst longer fragments come from heterochromatin regions which are cleaved by nucleases at lower efficiency (figure 1-2 and figure 2-1E). Prior to loading on gel, the purified DNA was treated again with RNAseA (Sigma Aldrich, USA) at a final concentration of 100µg/ml at 37ºC for 1 hr. 50-300bp fragments were isolated and purified from the gel using a minielute gel extraction kit (Qiagen, USA) as per manufacturer's instructions and validated by qPCR as before (Figure 2-1 B-C). Following this, the size selected sample was validated again by RT-PCR, this time using shorter amplicons to enable detection of the shorter fragments enriched by the size selection process.


A
DNase1 (u/ml)
20 30 35 40 45 50 60 70 80
10 kb
1 kb ladder


B
Pre cut
DNase1 (u/ml) 60 70 80
100bp ladder
300bp
Post Cut
DNase1 (u/ml) 60 70 80


C
80
70
60
50
40
30
20
10
0
60u/ml
70u/ml
80u/ml
β-actin/Chr18
TBP/β-actin
TBP/Chr18


D
Pre cut
DNase1 (u/ml) 70 80
Post Cut
DNase1 (u/ml) 70 80
300bp

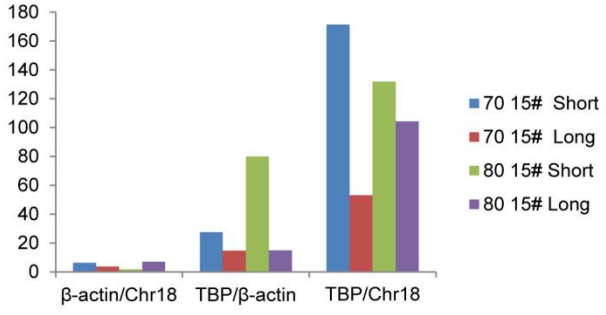

E
180
160
140
120
100
80
60
40
20
0
70 15# Short
70 15# Long
80 15# Short
80 15# Long
β-actin/Chr18
TBP/β-actin
TBP/Chr18

Figure 2-1 DNaseI treatment, size selection and NGS library preparation
DHS mapping in untreated SKH-1.

A) 0.7% agarose TAE gel, visualised with ethidium bromide, with DNaseI treated DNA run alongside a 1kb ladder (New England Biosciences, USA) with the 10kb marker labelled. Each well was loaded with a digest made with a different concentration of DNaseI. DNaseI concentrations used for each digest, increasing from 20u/ml to 80u/ml, are denoted above the gel.

B) Size selection of DNaseI treated DNA. Three digests made with different concentrations of DNaseI (60, 70 or 80u/ml) were loaded in the gel. 100bp ladder (New England Biosciences, USA) with the 300bp marker labelled. 50-350bp fragments excised from the 1.2% agarose TAE gel.

C) Validation by qPCR with amplicons corresponding to TBP promoter, β-actin gene body and chromosome 18 (negative control). Fragments from 60, 70 or 80u/ml DNaseI digested DNA were isolated by gel purification and used as templates for the qPCR experiments. Vertical axis denotes ratio of β-actin/chromosome 18, TBP/ β-actin or TBP/Chromosome 18.

D) Library for NGS after 15 cycles of amplification loaded in wells in a 2% agarose TAE gel: excision of short 190-250bp fragments (sequenced) and longer fragments 250-350bp (not sequenced). Input for library synthesis were size selected fragments from either 70 or 80u/ml digested DNA. Run alongside the samples was a 100bp ladder (New England Biosciences, USA) with the 300bp marker labelled.

E) Final qPCR validation of NGS library. Amplified, index-ligated material was gel purified as shown in D) and was used as a template for the qPCR validation. Results for “short’ fragments (190-250bp) alongside “long” fragments (250-350bp) for comparison. Short fragments are enriched from DHSs as nucleases cleave accessible DNA more efficiently than, inaccessible DNA (chromosome 18) or less accessible DNA (β-actin gene body). Hence, the ratio of both TBP/ β-actin and TBP/Chromosome 18 are higher in the shorter fragments than in the longer fragments.

2.11.1 Library production of DNaseI material for high throughput sequencing

A library suitable for high-throughput sequencing was prepared for untreated SKH-1 as follows. The sample was purified by using a minielute PCR clean up kit (Qiagen, USA) in between the following steps. All the enzymes used in library preparation were from New England Biolabs, USA. Overhanging strands of DNA were repaired by Klenow large fragment polymerase, T4 polynucleotide

kinase and T4 DNA polymerase in ligase buffer with deoxynucleotides. PCR at 20 ºC for 30mins was undertaken. Adenosine (0.2mM deoxyadenosine nucleotide) was added to 3' ends by incubation with Klenow fragment (5'-3' polymerase with 3'-5' exonuclease only, in order to fill in blunt ends), in NEB Buffer 2, at 37 ºC for 30mins. Illumina single read adaptors (part #1000521) used at 1/25 dilution were ligated to the fragments with T4 DNA ligase. Amplification of the library was performed by PCR using Phusion hot start polymerase with adaptor specific primers: 98°C for 30s, then 15 cycles of 98°C for 10s, 65°C for 10s, 72°C for 30s, finished with 72°C for 5 minutes. This material was run on a 2% agarose gel and 180-250bp fragments were isolated using the Minielute gel extraction kit as above. By concentrating shorter, sub-nucleosomal fragments into the NGS library, the number of cleavage fragments from within DHSs which are sequenced are increased (Figure 1-2, 2-1 D-E). At this stage the fragment size includes the adaptor sequences, 65bp in length, which is required to de-multiplex reads and to attach to the flow cell for cluster generation.

Purified t(3;21) patient CD34+ blasts (t(3;21) patient 1) were treated with DNaseI. After size selection, a library was prepared using Tru-seq DNA sample preparation kit (Illumina, USA) as per manufacturer's protocol, except after adaptor ligation, one round of Ampure bead purification (Beckman Coulter, USA) was performed instead of two and no purification after PCR was performed before size selection of the 180-250bp fraction by agarose gel . Ampure bead purification was performed by mixing a ratio of 1.8:1 beads to reaction volume, then the beads were washed twice with 70% ethanol and the DNA was eluted off the beads with 50 µl water.

Finally, size selected DNaseI treated material from t(3;21) patient 2 and SKH-1 transfected with RUNX1-EVI-1 or control siRNA was prepared for NGS using Microplex library preparation kit (Diagenode, Belgium). 14 cycles of PCR were used for t(3;21) patient 2 and 10-12 cycles of PCR for siRNA transfected SKH-1 cells. A final size selection step was performed by running the library on 1.5% TAE gel, followed by excision of 190-250bp sized gel fragment.

The quality of the libraries was assessed on an Agilent 2100 bioanalyser. Libraries were subsequently run on two lanes of an Illumina Hiseq flow-cell for transcription factor footprinting, or as part of 12 indexed libraries in one lane of a NextSeq500 (Illumina, USA) for DHS mapping alone.

2.12 ChIP-qPCR and ChIP-seq library preparation

2.12.1 Double cross-linking

A double cross-linking technique was used to optimise the efficiency of transcription factor chromatin immunoprecipitation (ChIP). 2 x 10^7 cells were washed thrice in PBS. Di(N-succinimidyl) glutarate (DSG) (Sigma-Aldrich, Germany) at 850µg/ml was added to 2x10^6 cells per ml and were incubated for forty-five minutes. Cells were washed four times and fixed with 1% formaldehyde (Pierce, Thermos Scientific, USA) for ten minutes. Glycine to produce a final concentration of 100mM was added to stop the reaction. The pellet was washed again with PBS. Buffer A (HEPES pH 7.9 10mM, EDTA 10mM, EGTA 0.5mM, Triton x100 0.25%, complete mini protease inhibitor cocktail (PIC) 1x (Sigma-Aldrich, Germany) was added for 10mins at 4°C and removed by centrifugation at 500g for 5 minutes. This was repeated with buffer B (HEPES pH 7.9 10mM, EDTA 1mM, EGTA 0.5mM, Triton x100 0.01%, PIC 1x). The residual nuclei were then spun down at 16000xg at 4°C for 5 minutes and aliquoted at 2x10^7 cells for 4 immunoprecipitations.

2.12.2 Chromatin immunoprecipitation (ChIP)

Each aliquot of 2x10^7 cells was resuspended in 600µl of sonication buffer (Tris-HCL pH8 25mM, NaCl 150mM, EDTA 2mM, Triton 100x 1%, SDS 0.25%, PIC 1x). 300µl of nuclei in sonication buffer was placed in each polystyrene tube and sonicated at 75% amplitude, 26 cycles: 30s on and 30s off per cycle (Q800, Active Motif, USA). Subsequently, 1.2ml of dilution buffer (Tris-HCL pH8 25mM, NaCl 150mM, EDTA 2mM, Triton 100x 1%, glycerol 7.5%, PIC 1x) was

added to the pooled post sonication material. This was divided equally between four immunoprecipitations (with 5% of input taken for validation).

15µl protein G beads (Diagenode, Belgium) were washed twice with 500µl of 50mM citrate phosphate buffer and once with 100mM sodium phosphate). 2µg antibody (EVI-1, C50E12, Cell signalling, lot 3; or RUNX1, Ab23980, Abcam lot 144722) or 4µg antibody (C/EBPα, A2814 Santa Cruz) was added to 10 µl 100mM sodium phosphate, 0.5% BSA and incubated with protein G beads at 4°C for 1 hour. Chromatin was then added to the protein G beads with antibody and returned to 4°C for 4 hours. Unbound chromatin was separated from the beads by magnet and the attached beads were washed by buffer 1 (Tris HCL 20mM, NaCl 150mM, EDTA 2mM, Triton x100 1%, SDS 0.1%), twice with buffer 2 (Tris HCL 20mM, NaCl 500mM, EDTA 2mM, Triton x100 1%, SDS 0.1%), LiCl buffer (Tris HCL 10mM, LiCl 250mM, EDTA 1mM, NP40 0.5%, sodium deoxychlolate 0.5%) and finally twice with wash buffer 4 (Tris HCL pH8, 10mM, NaCl 50mM, EDTA 1mM). The column was eluted twice with 50 µl buffer ($NaHCO_3$ 100mM and SDS 1%) and the eluant containing the chromatin was pooled. Crosslinks were reversed by incubating the samples at 65°C overnight in 500mM NaCl, 500 µg/ml proteinase K. DNA was purified by Ampure beads (Beckman Coulter, USA), as above, with the DNA eluted off with 50µl water. Validation of the ChIP was performed by qPCR using a standard curve of genomic DNA from untreated SKH-1 cells (10ng/µl followed by serial 1:5 dilutions). The input material was diluted 1:5 with water and qPCR was performed as above with primers listed in appendix. Validation was analysed as a ratio of the qPCR signal from the ChIP material over the input.

2.12.3 Library production of ChIP material for high throughput sequencing

Libraries for high throughput sequencing were prepared using the Tru-seq DNA sample preparation kit (Illumina, USA) or Kapa HyperPrep kit (Kapa biosystems, USA), as per manufacturer's protocol. 18 cycles of PCR was performed and 200-350bp fragments were size selected by running the samples in an agarose gel. Libraries were validated by qPCR, with an analysis of the

ChIP signal of a positive control region (e.g. PU.1 3H enhancer) over a negative control region (e.g. *IVL*). Finally, libraries were quantified by Kapa library quantification kit (Kapa biosystems, USA) and run in a pool of four indexed libraries in one lane of a HiSeq 2500 (Illumina, USA) or 12 indexed libraries in one lane of a NextSeq 500 (Illumina, USA) using 50 cycle single-end reads.

2.13 Cloning of RUNX1-EVI-1 into pSiew and LeGO vectors

The pME18s-RUNX1-EVI-1 plasmid was a gift of Dr K Mitani (Dokkyo Medical University, Japan). RUNX1-EVI-1 was amplified from the pME18s-RUNX1-EVI-1 plasmid by PCR; 10ng of template and Phusion polymerase (NEB, USA) with 3% DMSO. Amplification conditions were 98 °C 30s, followed by 5 cycles of 98°C 10s, 60°C 30s and 68°C 150s, followed by 25 cycles 98°C 10s, 68°C 30s and 72°C 150s, finished with 72°C 10mins. Primers incorporated a Bam-HI (5') and NotI (3') site. The PCR product was purified by Ampure beads (Beckman Coulter, USA) as above, except material was eluted off with 40µl water. 1 µg of purified PCR product was digested with 10 units of NotI and BamHI. The post digestion product was cleaned again by Ampure beads and was ligated to an empty LeGO-IG vector (gift from Boris Fehse, Hamburg, Germany) by T4 DNA ligase (NEB, USA) in a 3:1 digest to vector ratio. Ligated vector was used to transform chemically competent TOP10 E.Coli (Life Technologies, UK) which was spread onto ampicillin containing agar. DNA from picked colonies was purified with a silica-based column (Qiagen, USA) and inserts were confirmed by Sanger sequencing.

To clone RUNX1-EVI-1 into the pSIEW vector, RUNX1-EVI-1 was amplified from the LeGO-IG vector using the PCR conditions as above, but this time with primers that incorporated an AscI (5') and Bam-HI (3') site. The PCR product was purified by Ampure beads as above. 10 units of AscI and Bam-HI were used to digest the empty pSIEW vector and PCR product at 37ºC for 1 hour. The post digestion product was cleaned again by Ampure beads and was ligated to an empty vector by T4 DNA ligase (NEB, USA) in a 3:1 digest to vector ratio for 10 minutes at room temperature. Ligated vector was used to

transform chemically competent TOP10 E.Coli (Life Technologies, UK) which was spread onto ampicillin containing agar. DNA from picked colonies was purified with a silica-based column (Qiagen, USA) and inserts were confirmed by Sanger sequencing.

pSIEW empty vector, pSIEW RUNX1-ETO and pSIEW DNCEBP vector was a gift from Olaf Heidenreich (Newcastle, UK) (Bomken et al., 2013). The DNCEBP insert was originally developed by Charles Vinson (NIH, USA) (Krylov et al., 1995). The CEBPA-ER plasmid was a gift of Thomas Graf (Barcelona, Spain) (Bussmann et al., 2009). Schematic of regions flanked by LTR in the viral vectors shown in figure 2.

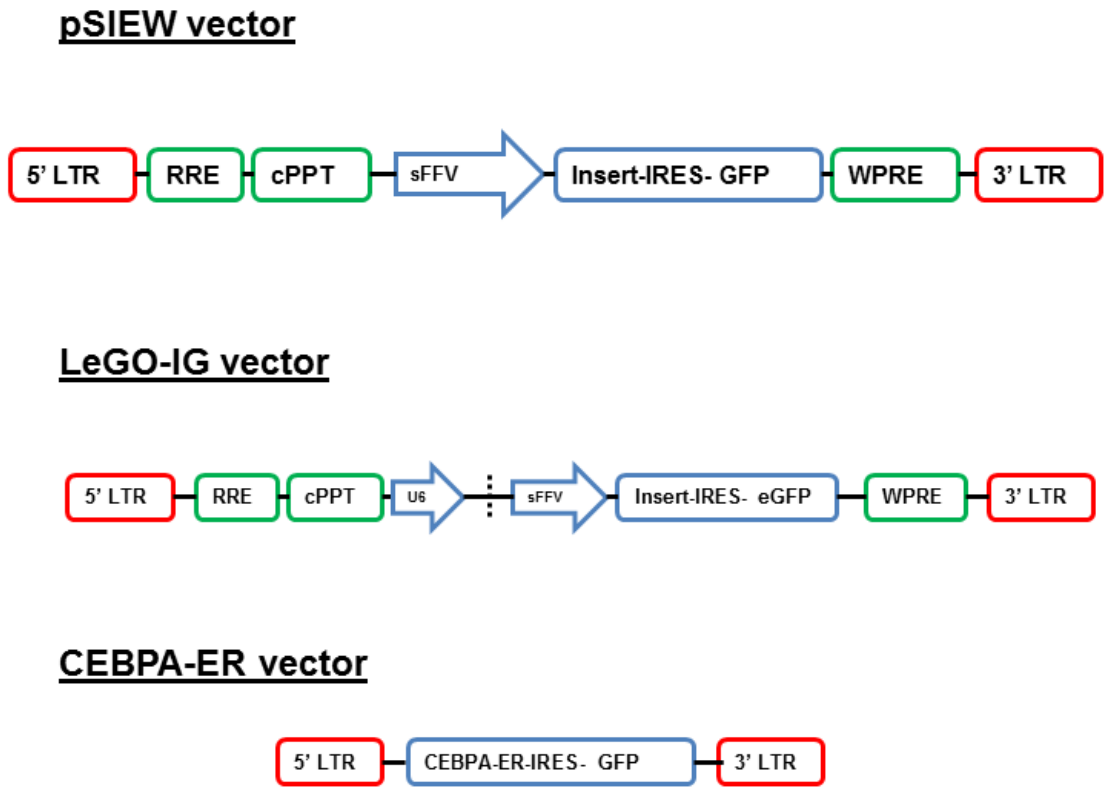


Figure 2-2 Schematic of viral vectors
LTR-long terminal repeats: enables integration of intervening sequence into host genome. cPPT: central polypurine tract recognition site for pro-viral DNA synthesis, increases transduction efficiency. WPRE: Woodchuck hepatitis virus post-transcriptional regulatory element increases nuclear export of transgene. IRES: internal ribosomal entry site allows synthesis of GFP. sFFV: spleen focus forming virus promoter. U6 promoter: allows cloning of shRNA if required.

2.14 Retroviral production

Lentiviruses are a part of the retrovirus family, and in this project pSIEW vector backbone (Empty, RUNX1-ETO, RUNX1-EVI-1 and DNCEBP) produces lentiviral particles. Packaging and envelope genes are on a separate plasmid to prevent further virus particle generation once transduced into the target cell. In

this project we also used CEBPA-ER plasmid to produce γ-retrovirus particles, and these require different viral packaging plasmids and transduction method as will be described below.

2.14.1 Transfection of HEK293T cells for lentiviral production

HEK293T cells were re-plated 24 hour prior to transfection, so that at time of transfection they were 80-90% confluent. On the day of transfection TransIT-293T (Mirius, USA) was brought to room temperature. Trans-IT-293T forms a complex with DNA plasmids to enable transfection into cells. A DNA mix was made from plasmids (Backbone vector containing transgene 30µg, and the packaging vectors: Tat 1.2µg, Rev 1.2µg, Gag/Pol 1.2µg, VSV-G 2.4µg (gift from George Murphy, Boston, USA) (Sommer et al., 2009). For each 15cm^2 2ml of Optimem serum free media was mixed with 90µl of TransIT-293 (Mirius, USA). This was allowed to rest at room temperature for 15 minutes. DNA mix was added to the TransIT-293 mixture and was left at room temperature for a further 15 minutes. Fresh DMEM with 10% FCS supplemented with glutamine and penicillin/streptomycin was exchanged with previous media on the HEK293T plates. The TransIT-293 – DNA mixture was then added dropwise to the HEK293T plate. Viral supernatant was collected after 48 hours and subsequently every 12 hours for 36 hours.

2.14.2 Transfection of HEK293T cells for CEBPA-ER virus production

HEK293T cells and TransIT-293T (Mirius, USA) was prepared as above. DNA mix was made from plasmids (Backbone vector containing transgene 36µg, Tat 1.2µg, Gag/Pol 30µg, Env 9µg (gift from James Mulloy, Cincinnati, USA) and was sufficient for 3x 10cm^2 dish. For each 10cm^2 dish 1.5ml of Optimem serum free media was mixed with 75µl of TransIT-293. This was allowed to rest at room temperature for 15 minutes. DNA mix was added to the TransIT-293 mixture and was left at room temperature for a further 15 minutes. Fresh DMEM with 10% FCS supplemented with glutamine and penicillin/streptomycin was exchanged with previous media on the HEK293T plates. The TransIT-293

– DNA mixture was then added dropwise to the HEK293T plate. Viral supernatant was collected after 48 hours and subsequently every 12 hours for 36 hours.

2.14.3 Virus concentration

The virus concentration technique was the same for all viral particles. Viral supernatant was centrifuged at 1660xg 4°C 15 minutes to pellet cell debris. The supernatant was then filtered through a 0.45μm disc filter.

The viral supernatant was concentrated using ultracentrifugation 120000xg for 2 hours 4°C in a SW28 rotor in an Optima^{TM} 100k-XL centrifuge (Beckman Coulter, USA). The supernatant was poured off and the viral pellet was re-suspended with the residual fluid. Alternatively the viral supernatant was concentrated using a Centricon Plus-70 100kDa filter (Millipore, USA), using the manufacturer's instruction. The column was pre-rinsed with sterile water and the column centrifuged at 2000xg, 25 minutes at 4 degrees. The column was then inverted and the concentrate recovered by centrifugation at 1000xg for 2 minutes.

2.14.4 Titration of viruses on HEK293T cells

Viruses were titrated before transduction of CD34+ PBSCs. 1x10^5 HEK293T cells were plated in each of 24 well plates 24 hours prior to transduction. On day of transduction viral concentrate was serially diluted 1:10 with media and media in the wells replaced with 450μl of fresh media. 50μl of the concentrate and subsequent 1:10 serial dilutions was added to the HEK293T cells. Polybrene at a final concentration of 8μg/ml was also added prior to spinoculation at 1500xg for 2 hours at 32°C. The principle of spinoculation is that through centrifugal force, viral particles and the cells due for transduction co-localise, thereby increasing the odds of viral entry into the cells. The plate was subsequently returned to the incubator overnight and at the next morning the viral media was removed and exchanged with fresh media. Viral transduction was estimated by eGFP percentage by flow cytometry 5 days after

viral transduction. Cells were washed with PBS and fixed with 1% formaldehyde for 10 minutes at room temperature. After washing again, the cells were detached by washing with PBS and taken to the flow cytometer. Dilution points with 1-20% eGFP positive cells were used to estimate the viral titre. At least an average of two dilutions is required for an accurate estimate.

eGFP-titre formula: Transducing Units/ml = (F × N)/ (100 × V × Df),

F=percentage of eGFP positive cells (in %)

N=number of cells spinoculated ($1x10^5$ in this example)

V= Volume of diluted viruses added to each well (50μl in this example)

Df=dilution factor of the viral concentrate at to each wells (x1 for the first well in this example)

2.14.5 Lentiviral transduction of CD34+ PBSCs and SKH-1

Cryopreserved CD34+ PBSCs were thawed with warmed IMDM with 20% BIT9500 serum substitute. The CD34+ PBSCs were incubated overnight in IMDM with serum substitute with cytokines and SR1 mix as described above in section 2.4.

CD34+ were transduced with viral concentrates (at a MOI of up to 36) with polybrene at 4μg/ml by spinoculation at 1500xg for 2 hours at 32°C in non-tissue culture treated plates. The plate was subsequently returned to the incubator overnight and at the next morning the viral media was removed and exchanged with fresh media. Viral transduction was estimated by eGFP percentage by flow cytometry 5 days after viral transduction. Cell sorting by FACS was undertaken by staining with either CD34-PE or CD34-APC (both Miltenyl Biotech, UK) and corresponding isotype control.

SKH-1 cells were transduced with lenitviruses (pSIEW empty vector or pSIEW DNCEBP) in the same manner, except polybrene was used at 8μg/ml. Cell sorting for GFP positive cells was undertaken by FACS.

2.14.6 Retroviral transduction with Retronectin

CEBPA-ER vector required RectroNectin (recombinant human fibronectin fragment, Takara, Japan) assisted transduction. RectroNectin has two domains that allow binding to a target cell expressing the integrin VLA-4 or VLA-5, but also has a third domain that enables binding to viral particles ((http://www.clontech.com/GB/Products/Viral_Transduction/Hematopoietic_Cell_Transduction/ibcGetAttachment.jsp?cItemId=61561&fileId=6521143&sitex=10030:22372:US), accessed 3rd September, 2016). Non-tissue culture treated 6 well plate was incubated with 2ml RetroNectin solution in PBS at 24µg/ml for 2 hours at room temperature. The RectroNectin solution was subsequently removed and the plate blocked with 2ml PBS with 2% BSA for 20minutes at room temperature. Finally, the PBS/BSA solution was removed and the wells washed with 2ml Hank Balanced Salt solution containing 2.5% HEPES.

Concentrated virus was coated onto the wells by centrifugation for 45 minutes at 2000xg at room temperature. Virus loading was repeated once more before $1x10^6$ SKH-1 cells, in media mixed with concentrated virus (50:50 ratio), at a final concentration of $0.5x10^6$ /ml was added to the wells with polybrene at 8µg/ml. The plate was left overnight at 37ºC, 5% CO_2 in a humidified incubator. On the next day, the cells were centrifuged at 300xg at room temperature for 5 minutes and the cells were re-suspended in standard RPMI 1640 with 10% FCS, supplemented with penicillin, streptomycin and glutamine. CEBPA-ER transduced SKH-1 cells were isolated by FACS using GFP labelling.

2.15 Titrating 17 β-estradiol treatment of CEBPA-ER SKH-1 cells

CEBPA-ER plasmid encodes a C/EBPα protein fused to an oestrogen receptor ligand binding domain. Upon binding to estradiol, the fusion protein can translocate into the nucleus (Umek et al., 1991). We tested different concentrations of 17 β estradiol (Sigma Aldrich, Germany), after 48 hour exposure, on CEBPA-ER SKH-1, using expression of CD11b APC by flow cytometry as a read-out.

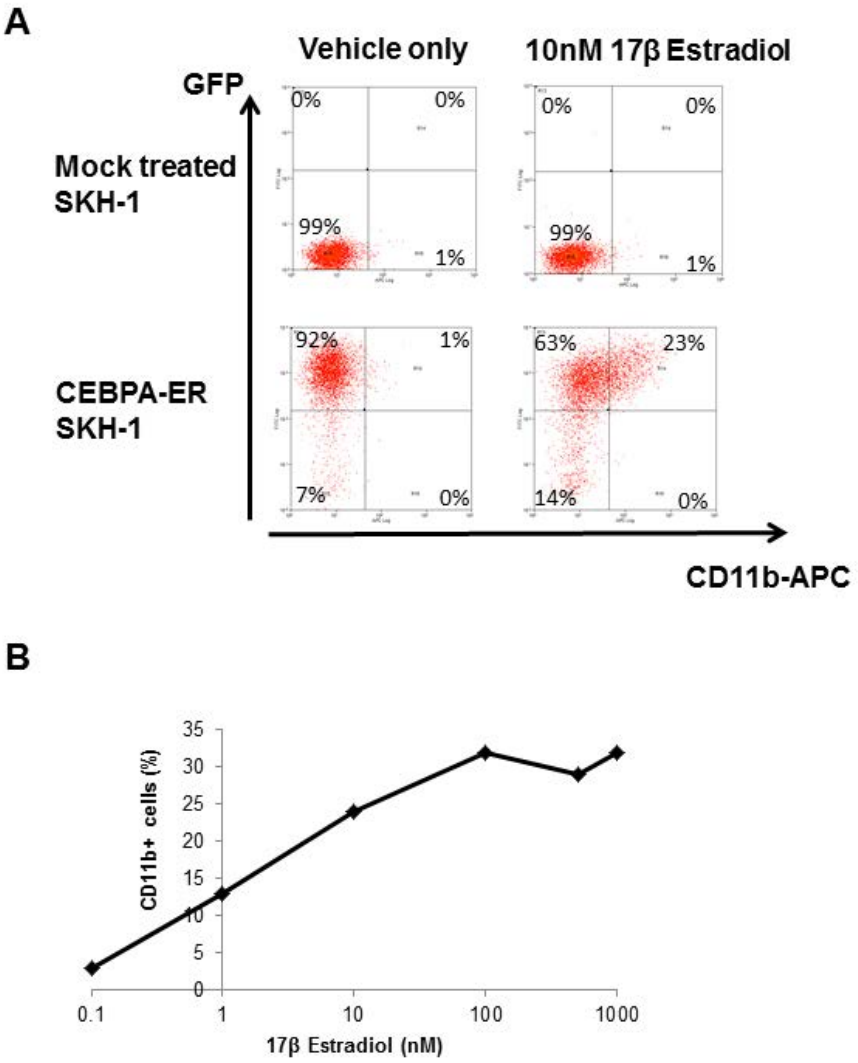


Figure 2-3 Titration of 17β-Estradiol on CEBPA-ER SKH-1 cells
Flow cytometry of mock treated or CEBPA-ER transduced SKH-1 cells treated for 48 hours with either vehicle only (ethanol) or 17β-Estradiol. Cells stained with CD11b-APC. CEBPA-ER vector contains a sequence encoding GFP.

A) Flow cytometry plot showing percentage of SKH1 cells in each quadrant B) Graph showing CD11b positive percentage of total cells following titration of doses of 17β-Estradiol after 48 hours.

2.16 Methylcellulose colony forming culture assay of CEBPA-ER cells.

Mock transduced and CEBPA-ER transduced SKH-1 cells were kept at $0.25x10^6$ cells per ml and treated with either 10nM 17β-Estradiol or equivalent volume of ethanol for 48 hours.

IMDM with 2.6% methylcellulose premixed was obtained from Stem Cell Technologies, USA (H4100 methocult). To 1.2ml of pre-mixed media, a further 1.2ml of IMDM, and 600µl of heat-inactivated FCS were added, HEPES (Sigma Aldrich, Germany) was also added for a final concentration of 25mM. 17β-estradiol was added for a final concentration of 10nM, and in the comparative assay an equal volume of ethanol was added.

To each 3ml mixture of methylcellulose, IMDM and FCS 15000 cells were added and were shaken vigorously. 1.1ml of this mix of cells and methylcellulose media was added to a 35mm non-tissue culture treated dish, in duplicate. These small dishes were subsequently placed in a larger 150mm plate humidified by a well in the centre containing sterile water. These dishes were kept in an incubator at 37°C and 5% CO_2. Colonies of more than 20 cells were counted after 7-10 days and for each experiment an average was taken of the two replicates.

2.17 Antibody staining for flow cytometry

$15x10^4$ were centrifuged at 300xg and washed with MACS buffer. The cell pellet was resuspended in 50µl MACS buffer and 2µl of antibody was added and incubated for 15minutes at 4ºC in the dark. After incubation, the cells was washed once with MACS buffer before resuspension in 300µl MACS buffer and analysed on Cyan ADP (Beckman Coulter, USA). Data was analysed on Summit 4.3 (Beckman Coulter, USA). Antibodies used in this project are listed below.

2.18 Whole cell lysate preparation by RIPA buffer lysis

Whole cell lysate was made by lysing $5x10^6$ cells using RIPA buffer (50 mM Tris pH7.4, 150 mM NaCl, 0.25% Na-deoxycholate, 1% NP40, 1mM PMSF, complete mini protease inhibitor cocktail (PIC) 1x (Sigma-Aldrich, Germany)). After incubation on ice for 15 minutes, the sample was sonicated for 1 min using a Bioruptor (Diagenode, Belgium) at 4°C.

2.18.1 Quantification by Bradford reagent

Protein extracts were quantified using Bradford protein reagent (Bio-Rad, USA) and 595nM absorbance quantified by spectrophotometry. Absolute concentrations were determined using a standard curve from a known concentration of BSA (Pierce, USA).

2.19 Nuclear extract

Nuclear extracts were prepared using a co-immunoprecipitation kit (Active Motif, USA).

2.20 Western blotting

Cell lysate or nuclear extracts was run on an acrylamide gel and transferred to nitrocellulose membrane. This was probed with the antibodies described in the appendix. Enhanced chemiluminescence by SuperSignal PICO (Thermos Scientific, USA) was used to develop the membrane. Chemiluminescence was detected using either developer or Chemidoc XRS system (BioRad, USA).

2.21 Lipofectamine transfection of RUNX1-EVI-1 plasmid

HEK293T and HeLa cells were transfected by lipofectamine (Life Technologies, UK) as per manufacturer's protocol.

DATA ANALYSIS

2.22 Analysis of *CEBPA* expression in large cohorts of patients with AML

RSEM normalised count from either normal karyotype AML (n=63) vs CBF fusion protein AML (t(8;21): n=7, t(3;21): n =1) from the TCGA project (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/# accessed 30th June 2016, dataset labelled "TCGA acute myeloid leukemia (LAML) gene expression by RNA-seq (Illumina HiSeq)'). Normalised signal from gene expression microarray transformed by Log2 from Verhaak et al., 2009 (Verhaak et al., 2009) was also analysed. Normal karyotype AML (n=187) vs CBF fusion protein AML (t(8;21): n=35, t(3;21): n =1). Data downloaded from GEO, accession number GSE6891, accessed 30th July 2016. Boxplot generated by SPSS v21 (IBM, USA). Unpaired T-test used to compare gene expression between groups.

2.23 ChIP and DNaseI sequencing data Analysis

2.23.1 Alignment

Sequences from all ChIP and DNaseI sequencing experiments in fastq format were mapped onto the reference human genome version hg38, Genome Reference Consortium GRCh38. The quality control statistics for the samples were obtained using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw reads were aligned to the reference genome using Bowtie2 (Langmead and Salzberg, 2012). Reads from ChIP-seq data that were uniquely aligned to chromosomal positions were retained and duplicate reads were removed from the aligned data using Picard tools (http://broadinstitute.github.io/picard/). The aligned reads were used to generate density profiles using "genomeCoverageBed" function from bedtools (http://bedtools.readthedocs.org/en/latest/). These tag densities were displayed using the UCSC Genome Browser (Kent et al., 2002). The numbers of aligned reads are listed in Table S7. RUNX/ETO ChIP is the

pool of publically available data downloaded from GEO with accession numbers GSM1113429**,** GSM1113430 (Ben-Ami, et. al, 2013) and GSM1082306 (Wang et. al, 2013).

2.23.2 Peak calling

Regions of enrichment (peaks) of ChIP and DNAse1 sequencing data were identified using DFilter software (Kumar et al., 2013) with recommended parameters (-bs=100 -ks=50 –refine). Peak overlaps, gene annotations were performed using in-house scripts. High confidence ChIP-Seq peaks were defined as those overlapping peaks in the DNaseI-seq data. Overlaps between ChIP and DNaseI sequencing were defined by requiring the summit of a peak in the ChIP dataset to lie between start and end coordinates of a peak in the DNaseI data. Peaks were allocated to genes if located in either their promoters or within the region of 2000 bp downstream and 2000 bp upstream of the transcription start sites (TSS), as intragenic if not in the promoter but within the gene body region, or if intergenic, to the nearest gene located within 100 kb. Overlaps between ChIP-seq peaks were defined by requiring the summits of two peaks to lie within +/-200 bp.

2.23.3 Clustering of ChIP and DNaseI sequencing data

Hierarchical clustering with Euclidean distance and complete linkage clustering was used for clustering of transcription factors based on similar binding patterns of different ChIP-seq data, in SKH-1 cells. The high confidence peaks for all transcription factors were intersected and merged when overlapping. The read counts for all union peaks were normalised with regards to total reads depth counts and then Pearson's correlation coefficients were calculated between samples using log2 of the normalised read counts. A correlation matrix was generated and Pearson correlation coefficients are displayed after hierarchical clustering as a heatmap. Colors in the heatmap indicate the strength of association between each pair of transcription factors. Heat maps were

generated using Mev from TM4 microarray software suite (Saeed et al., 2006). Same way was used for DNase-seq data clustering.

2.23.4 Average tag density profile and heatmap

The tag density and average profiles were generated by calculating the tag density normalised as coverage per million reads within 4kb of the DNAse1 peak summit. The read counts for all union peaks that were computed. Coverages were calculated for all union peaks and ranked by log2 fold change. Heatmap images were generated via Java TreeView (http://jtreeview.sourceforge.net/) and average profiles were plotted using R (https://www.r-project.org/).

2.23.5 Motif identification and clustering

De novo motif analysis was performed on peaks using HOMER (Benner et.al 2010). Motif lengths of 6, 8, 10, and 12 bp were identified in within ± 200 bp from the peak summit. The top enriched motifs with a significant p value score were recorded. The annotatePeaks function in HOMER was used to find occurrences of motifs in peaks. In this case we used known motif position weight matrices (PWM) from HOMER database.

2.23.6 Motif clustering

Digital footprinting of t(3;21) AML patients 1 and 2 and t(8;21) AML patients 1 and 2 from DNaseI high-depth sequencing data was performed using the Wellington algorithm (Piper et al., 2013) with FDR=0.01. For the heatmap that shows hierarchical clustering of motif occurrences within RUNX1/EVI1 footprints a motif positions search was done within peaks that are only footprinted in t(3;21) patients. The distance between the centres of each motif pairs was calculated and the motif frequency was counted if the first motif was within 50bps distance from the second motif. Z-scores were calculated from the mean and standard deviation of motif frequencies observed in random sets using bootstrap analysis. For bootstrapping, peak sets with a population equal to that of the footprinted peaks were randomly obtained from the union of t(3;21),

t(8;21) and CD34+ DNase-seq footprints. Motif search was repeated for each random set and then the mean and the standard deviation for the total motif frequencies of the random peak sets were calculated and compared with the actual motif frequencies to obtain the Z- scores. A matrix was generated and Z scores were displayed after hierarchical clustering as a heat map. Red colour means that motifs are overrepresented and grey colour indicates that motif is underrepresented. The same procedure was repeated with RUNX1/ETO and RUNX1 peaks that are only footprinted in t(8;21) patients and where motif search was done exactly within the footprint coordinates and the random sets were generated from the total patient's footprints.

2.23.7 Motif enrichment

To identify motifs (identified by HOMER) that are relatively enriched in the distal transcription factors (TFs) sites of one cell type compared to another or one TF compared to another from same cell type we considered all possible comparisons, these being TF sites in (A) which are not shared with each of the other TFs (B). For a given set *j* of TFs, we defined a motif enrichment score (S_{ij}) for motif *i* in peak set *j* as

$$S_{ij} \frac{n_{ij}/M_j}{\sum_j n_{ij}/\sum_j M_j}$$

where n_{ij} is the number of peaks in each subset *j (j=1,2,…,12)* containing motif *i (i=1, 2,….,I), I* is the total number of motifs used in the test, and M_j the total number of peaks in each subset *j (j=1,2,…,30).* A matrix was generated and the motif enrichment scores were displayed as a heatmap after hierarchical clustering with Euclidean distance and complete linkage. The heatmap was generated using Mev from TM4 microarray software suite (Saeed et al., 2006).

2.24 RNA-seq data Analysis

RNA-Seq reads were aligned to the hg38 human genome build using STAR. Separate density profiles for the positive and negative strand were generated

for RNA-seq data. Fragments per Kilobase of transcript per Million mapped reads (FPKM) values for each gene were extracted using Cufflinks and differentially expressed genes were extracted using the limma R package (Ritchie et., al. 2015), in this case the count data was used as an input to limma. All genes with p-value ≤ 0.01 were considered and at least 1.5-fold changes between before and after RUNX1/EVI1 knock down. The differentially expressed genes for the AML patients were considered with at least 2-fold changes using the CD34+ PBSC as a control.

The correlation between any two samples was obtained as the Pearson correlation coefficient of expression values over all genes. A correlation matrix was thus generated for all the samples and hierarchically clustered.

Clustering of gene expression was carried out on signal intensity for all expressed genes and on fold-changes for genes associated with at least a 1.5-fold change. Hierarchical clustering was used with Euclidean distance and average linkage clustering. Heatmaps were generated using Mev (Saeed et al., 2006).

The GSEA software (Subramanian et al., 2005) was used to perform gene set enrichment analysis on group of genes. The normalised enrichment score (NES), the p-value and the FDR q-value are displayed on the enrichment plot.

Gene ontology (GO) analysis was performed using Bingo (Maere et al., 2005) and David online tool at david.abcc.ncifcrf.gov (Huang da et al., 2009) using Hypergeometric for overrepresentation and Benjamini and Hochberg (FDR) correction for multiple testing corrections. KEGG Pathway network analysis was performed using clueGO tools (Bindea et al., 2009) with kappa score = 0.3. The right-sided enrichment (depletion) test based on the hypergeometric distribution was used for terms and groups. Groups were created by iterative merging of initially defined groups based on the kappa score threshold. The relationship between the selected terms is defined based on their shared genes and the final groups are randomly coloured where one, two colours or more represents that a gene/term is a member of one, two or more groups respectively. The size

of the nodes reflects the enrichment significance of the terms. The network is laid out using the layout algorithm supported by Cytoscape.

2.24.1 Gene ontology (GO) Clustering

Gene ontology (GO) analysis was perform for the up and down regulated genes that change expression 1.5 fold after RUNX1/EVI1 Knock down in SKH-1 cells and similar analysis was also perform for the up and down regulated genes that change expression after RUNX1/ETO Knock down in Kasumi-1 cells. GO terms from all four groups were combined and a matrix with GO terms enrichment p-values was generated then a hierarchical clustering was used with Euclidean distance and average linkage.

2.25 Accession Numbers

The data generated in this study were deposited at NCBI under accession number GSE87286.

Previous published data were downloaded from the Short Read Archive (accessions RUNX1 AML blast CD34+ ChIP-Seq: GSM1466000, CD34+ PBSC-1 DNAse1-Seq: GSM1466003, t(8;21) patient#1 DNAse1-seq: GSM1466005, t(8;21) patient#2 DNAse1-seq: GSM1466004, RNA-seq from Kasumi-1 cells: GSE54478, RUNX1 kasumi-1 cells: GSM850823). RUNX/ETO ChIP-Seq is the combination of publically available data downloaded from GEO with accession numbers GSM1113429, GSM1113430 and GSM1082306.

2.26 Tables of Primers and Antibodies

Table 2-2 Primers for RT-qPCR

Name	Forward	Reverse
GAPDH	CCTGGCCAAGGTCATCCAT	AGGGGCCATCCACAGTCTT
RUNX1-EVI-1	CCACAGAGCCATCAAAATCA	TCTGGCATTTCTTCCAAAGG
RUNX1	CCCTCAGCCTCAGAGTCAGAT	AGGCAATGGATCCCAGGTAT
CEBPA	GAGGGACCGGAGTTATGACA	AGACGCGCACATTCACATT
CEBPB	GACAAGCACAGCGACGAGTA	CTCCAGGTTGCGCATCTT

GATA2	CAGACGAAGGCAACCATTTT	GCTCAGACCACCAAGTCTCC
HOXA9	GTGATGCCATTTGGGCTTAT	GGGGTGAGAGAAGGGAGAAG
MEIS1	CAGAAAAAGCAGTTGGCACA	GGTCTATCATGGGCTGCACT
CSF1R	GCGGGACTATACCAATCTGC	AGCAGGTCAGGTGCTCACTA
MPO	CCAACAACATCGACATCTGG	GCTGAACACACCCTCGTTCT
CTSG	TCCTGGTGCGAGAAGACTTTG	GGTGTTTTCCCGTCTCTGGA
CEBPA-ER	GCTGGAGTTGACCAGTGACA	AAGGTTGGCAGCTCTCATGTC

Table 2-3 Primers for validating DNaseI

Target	Forward	Reverse
Actin (short)	GCAATGATCTGAGGAGGGAAGGG	AGCTGTCACATCCAGGGTCCTCA
TBP promoter (short)	CTGGCGGAAGTGACATTATCAA	CCCGACCTCACTGAACCC
Chromosome 18 (short)	AGGTCCCAGGACATATCCATT	GTTCAAATTGTGTTTTGTGGTTA
Actin (long)	GCAATGATCTGAGGAGGGAAGGG	GTGTCTTTCCTGCCTGAGCTGAC
TBP promoter (long)	CTGGCGGAAGTGACATTATCAA	GCCAGCGGAAGCGAAGTTA
Chromosome 18 (long)	ACTCCCCTTTCATGCTTCTG	AGGTCCCAGGACATATCCATT

Table 2-4 Primers for ChIP-qPCR

Target	Forward	Reverse
PU.1 3H Enh	AACAGGAAGCGCCCAGTCA	TGTGCGGTGCCTGTGGTAAT
IVL	GCCGTGCTTTGGAGTTCTTA	CCTCTGCTGCTGCCACTT
MPO	CAACACACTCACACCCCACT	TGGGAACTCTAAGTGGGCAG
CTSG	AGACCGTGTAATCCAAGCCA	TCTCGGCACTGACTTAGCAG
TREM1	ACAAGGCACCACAATGACCT	GGCCTCATATCCTGTTGTGC
SIGLEC1	GTATCAGGGGCTGCTTCCTC	CTGGGTTGGACAGTAGAGCT

Table 2-5 Primers of DNaseI PCR (figure 3-21 C)
Primers for MPO, TREM1 and SIGLEC1 as above

Target	Forward	Reverse
CTSG	GGTTTCATCACCCAAGGCTG	TGGCTTGGATTACACGGTCT

Table 2-6 Antibodies for probing Western blots

Antibody target	Company	Serial number
EVI-1	Cell Signalling	2593
RUNX1 (C-terminal epitope)	Abcam	23980
RUNX1 (N-terminal epitope)	Cell Signalling	4334
C/EBPα	Abcam	40761
FLAG epitope	Sigma	F3165
GAPDH	Abcam	8245
Anti-Rabbit HRP	Cell signalling	7074
Anti-Mouse HRP	Jackson	115 035 062

Table 2-7 Antibodies for ChIP

Antibody target	Company	Serial number
EVI-1	Cell Signalling	2593
RUNX1	Abcam	23980
C/EBPα	Santa Cruz	A2814

Table 2-8 Antibodies for Flow cytometry

Antibody target-flurochrome	Company	Serial number
CD34-APC	Miltenyl-biotech	10098139
CD34-PE	Miltenyl-biotech	130081002
CD117-APC (Clone A3C6E21)	Miltenyl-biotech	130091733
CD11b-APC	Miltenyl-biotech	130091241
CD11b PE (Clone M1/70)	Ebiosciences	120011281
CD14-FITC	Miltenyl-biotech	130080701
CD14-PE (Clone MΦP9)	BD Biosciences	562691
Annexin V-FITC/PI kit	BD Pharmingen	556547
Annexin V-APC/PI kit	Ebiosciences	88-8005-74
IgG FITC	Miltenyl-biotech	130093192
IgG PE	Miltenyl-biotech	130093193
IgG APC	Miltenyl-biotech	130093194

Chapter 3. Results

3.1 Epigenetic landscapes differ between t(3;21) and t(8;21) leukaemia

3.1.1 DNase-seq identifies distinct epigenetic features of t(3;21) as compared to t(8;21) leukaemia in patient CD34+ blasts

DHS are nuclease accessible regions that represent gene regulatory elements (Cockerill, 2011). We have previously shown that different pattern of DHSs is associated with *FLT3*-ITD mutant AML as compared to normal karyotype AML (Cauchy et al., 2015). This was proof of principle that distinct DHS signatures are identifiable and can be linked to specific AML subtypes. Given the clinical differences in patients with t(3;21) and t(8;21) AML we sought to identify the underlying differences in epigenetic landscape through genome-wide mapping of DHSs by DNase-seq. We purified CD34+ populations from patients with t(3;21) AML using cell-surface antibody labelling and compared this to DNase-seq data from t(8;21) CD34+ patient blasts that have been previously generated by the lab (Ptasinska et al., 2012, Ptasinska et al., 2014). As controls we used CD34+ purified G-CSF mobilised PBSCs from healthy donors undergoing apheresis for stem cell transplantation (sample details in table 2-1). All DNaseI-seq libraries were run at high read depth (50 cycles, single end), with alignment statistics and peaks identified shown in table 3-1. Corresponding RNA-seq libraries for all samples were run at 100 cycle, paired ends. The number of reads aligned for each experiment are shown in table 3-2. We also performed the same experiments on t(3;21) SKH-1 cell line expressing RUNX1-EVI-1 (Mitani et al., 1994).

Table 3-1 DNase-seq sequencing results
DNase-seq reads alignment and peaks detected in t(3;21) SKH-1 cell line, primary CD34+ AML blasts and normal CD34+ PBSC.

DNase-seq Dataset	Aligned reads (n)	Peaks (n)
CD34 J209 (PBSC 1)	173,261,382	36,041

CD34 R299 (PBSC 2)	205,284,104	36,088
t(3;21) SKH1 cell line	348,530,966	28,821
GT027 (t(3;21) patient 1)	407,335,267	31,532
AML5354 (t(3;21) patient 2)	237,196,453	35,666
H12812 (t(8;21) patient 1)	405,680,774	32,262
H18901 (t(8;21) patient 2)	387,658,545	35,052

Table 3-2 RNA-seq reads alignment
RNA-seq reads alignment in t(3;21) SKH-1 cell line, primary CD34+ AML blasts and normal CD34+ PBSC.

RNA-seq Dataset	Aligned reads (n)
t(3;21) SKH-1 cell line replicate 1	30334985
t(3;21) SKH-1 cell line replicate 2	11898204
CD34 R454 (PBSC 1)	16538775
CD34 R423 (PBSC 2)	12402996
H12812 (t(8;21) patient 1)	12347883
H18901 (t(8;21) patient 2)	11331201
GT027 (t(3;21) patient 1)	48166460
AML5354 (t(3;21) patient 2)	11969780

Reads were processed and aligned to the hg38 assembly. UCSC browser screenshots (Kent et al., 2002) demonstrate differences in gene expression and DHSs responsible for these differences between t(8;21) and t(3;21) leukaemia (figure 3-1). For example, *MEIS1* is a critical rate-limiting regulator of LSC potential in leukaemias induced by MLL fusion proteins (Wong et al., 2007). Here, *MEIS1* was expressed in t(3;21) but not in t(8;21) leukaemias and this correlated with the presence of a DHS in the promoter and a previously characterised distal enhancer (Xiang et al., 2014) in the t(3;21) but not t(8;21) samples. MEIS1 and HOXA9 directly interact with each other to immortalise myeloid progenitors (Schnabel et al., 2000). Both *HOXA9* and *HOXA10* were expressed in t(3;21) cells but not t(8;21) leukaemia (figure 3-1A). ROBO1 is a

cell surface receptor and is thought to be a tumour suppressor in MDS and AML (Xu et al., 2015). *ROBO1* and *CACNA2D2* were expressed in t(8;21) but not t(3;21) leukaemia (figure 3-1B). This fact was associated with differences in DHSs between each type of leukaemia at both these genes, suggesting these differences in gene expression between each CBF leukaemia are due to differences in activity of gene regulatory elements.

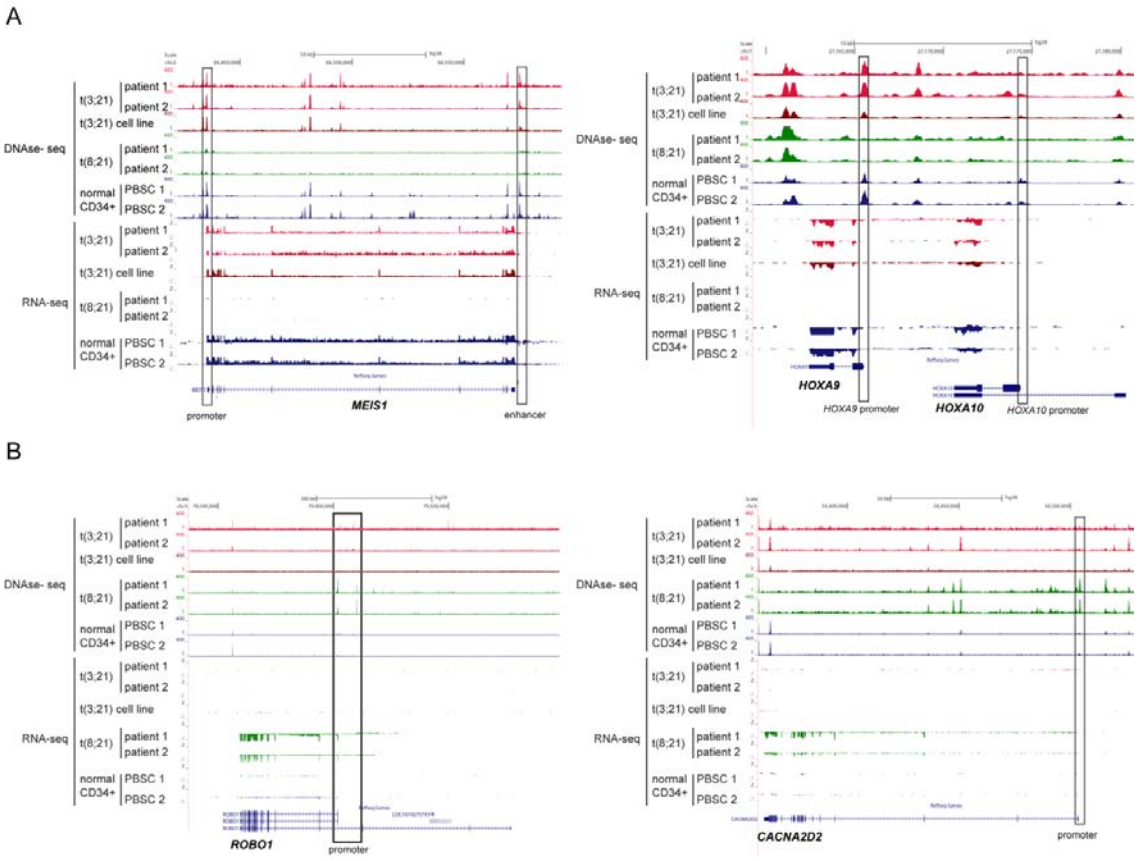


Figure 3-1 Epigenetic landscapes differ between CBF leukaemia which is associated with differences in gene expression.
UCSC browser screenshot of DNase-seq and RNA-seq aligned reads from two patients with t(3;21) AML, two patients with t(8;21) AML, t(3;21) cell line and normal CD34+ PBSC. *MEIS1*, *HOXA9* and *HOXA10* are expressed in t(3;21) but not t(8;21) cells. *ROBO1* and *CACNA2D2* are expressed in t(8;21) but not t(3;21) cells.

A) *MEIS1* locus showing the presence of a DHS at the promoter and distal enhancer +140kb from the promoter in t(3;21) cells but not in t(8;21) cells. DHS are present in the promoter of *HOXA9* and *HOXA10* loci in t(3;21) cells but not in t(8;21) cells.

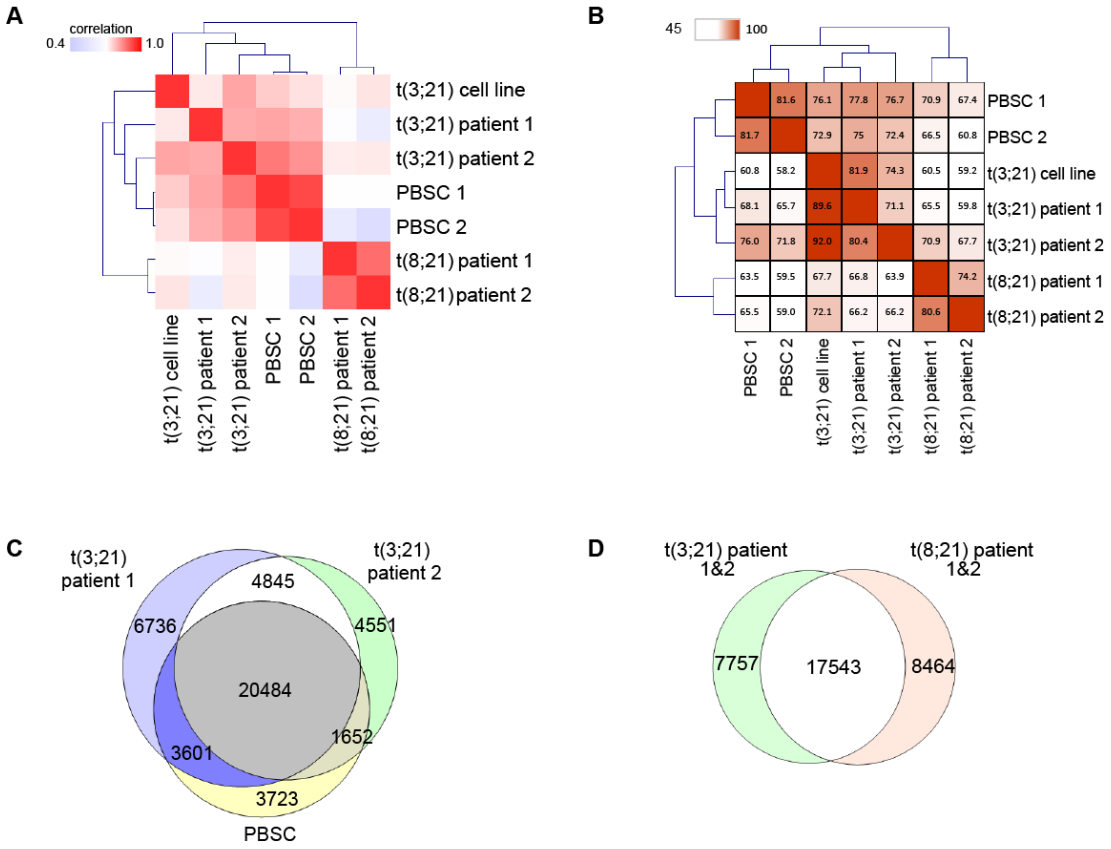
B) DHS are present at promoter and regulatory elements in *ROBO1* and *CACNA2D2* in t(8;21) but not in t(3;21) cells.

These examples of specific differences between t(3;21) and t(8;21) AML are part of a genome wide pattern of differences between the two types of AML. Figure 3-2A shows that in a hierarchical clustering of DNase-seq signals, t(8;21) leukaemias cluster separately to t(3;21) cells and normal PBSCs. A pairwise analysis of the overlap in DHSs show that t(8;21) leukaemia, t(3;21) leukaemia and the normal PBSCs form three major, separate, clusters (figure 3-2B). Notably, the t(3;21) cell line and patient blasts formed a central cluster because of the substantial overlap in DHSs. 90% of DHSs found in t(3;21) cell line were also found in t(3;21) patient CD34+ blasts. In contrast, approximately 70% of DHSs in t(3;21) cell line were found in either t(8;21) AML or normal PBSCs. This strong correlation between t(3;21) cell line and patient samples validates the t(3;21) cell line as a model for studying the epigenome of t(3;21) leukaemia.

Figure 3-2C shows substantial overlap in DHSs between each t(3;21) patient, but also a large overlap with normal PBSCs, demonstrating that they share a large number of gene regulatory elements. To identify the underlying drivers of the differences in DHSs between t(3;21) and t(8;21) leukaemia we determined DHSs found in leukaemia from both t(3;21) or both t(8;21) (figure 3-2D) and then performed de novo motif analyses on DHSs unique to each leukaemia type (figure 3-2E). We found GATA and IRF motifs were unique to t(3;21) specific DHSs, whilst E-box and CEBP motifs were unique to t(8;21) specific DHSs.

The identification of unique motifs can be used in association with the gene expression data to suggest which transcription factor might be specifically involved in the regulation of each leukaemia. For example, the enrichment of GATA motif containing DHSs in t(3;21) as compared to t(8;21) cells may be due to GATA2. *GATA2* is more highly expressed in t(3;21) leukaemia than in t(8;21) leukaemia (figure 3-7E). GATA2 is a critical regulator of HSC homeostasis: knockout of *Gata2* in mice prevents the formation of HSCs (Tsai et al., 1994) and over-expression of *Gata2* results in a block of haematopoietic differentiation

(Persons et al., 1999). The identification of E-box motifs unique to DHS specific for t(8;21) leukaemia as opposed to t(3;21) leukaemia are consistent with the recognised direct interaction of E-box transcription factors such as HEB and E2A with RUNX1-ETO (Sun et al., 2013, Ptasinska et al., 2014). As E-box transcription factors interact directly with the NHR domain (figure1-4) (Sun et al., 2013), this interaction is unique to RUNX1-ETO and not to RUNX1-EVI-1.



E

unique to t(3;21) not t(8;21)

Motif	match	score	% targets
	CTCF	1e-754	19.3
	ETS	1e-335	24.3
	AP1	1e-226	16.8
	IRF	1e-154	25.2
	RUNX	1e-54	14.4
	GATA	1e-36	11.9

unique to t(8;21) not t(3;21)

Motif	match	score	% targets
	ETS	1e-606	54.8
	RUNX	1e-212	24.2
	AP1	1e-144	15.8
	CTCF	1e-115	5.7
	C/EBP	1e-111	19.1
	Ebox	1e-89	57.7

Figure 3-2 DHS patterns segregate patients according to CBF fusion translocation

A) t(3;21) leukaemia segregates with CD34+ PBSC and away from t(8;21) leukaemia. DNase-seq data from two patients with t(3;21), two patients with t(8;21), two independent CD34+ PBSC and t(3;21) SKH-1 cell line were analysed. The matrix shows the correlation in DNase-seq data between the different samples. Clustering of samples is based on the strength of correlation between samples.

B) Three major clusters are comprised of either CD34+ PBSCs, t(3;21) leukaemia (both primary patient samples and cell line) and finally t(8;21)

leukaemia from primary patient material. Overlap of DHS peaks in a pairwise comparison between each sample (t(3;21) patient and cell line with t(8;21) patients and normal CD34+PBSC).

C) Large number of DHSs are shared between both t(3;21) primary patient CD34+ purified cells and healthy CD34+ PBSCs. Venn diagram showing peak overlaps between t(3;21) patient 1, t(3;21) patient 2 and normal CD34+ PBSC.

D) Identification of DHSs unique to either t(3;21) or t(8;21) leukaemia. Venn diagram showing DNase-seq peak overlaps between t(3;21) patients (both patients combined), and t(8;21) patients (both patients combined).

E) IRF and GATA motifs are unique to t(3;21) specific DHSs, whilst E-box and C/EBP motifs are unique to t(8;21) specific DHSs. De novo motif discovery in distal DHSs unique to t(3;21) as compared to t(8;21) patients, and in distal DHSs unique to t(8;21) compared to t(3;21) patients, as shown in D).

3.1.2 Gene expression differs between patients with different CBF leukaemia

Having demonstrated a distinct DHS signature in t(8;21) leukaemia as compared to t(3;21) leukaemia, we wanted to know whether this translated into differences in gene expression between the cells expressing the two forms of CBF fusion protein. To answer this question, we used unsupervised clustering techniques to group the same samples based on their gene expression profiles, as measured by RNA-seq. RNA-seq signals from t(8;21) CD34+ patient cells, t(3;21) CD34+ patient cells and normal PBSCs clustered independently (figure3-3A). Principle component analysis (PCA) provides another way of grouping this datasets in an unsupervised manner. PCA transforms the large number of RNA-seq measurements (dimensions) to components that represent the majority of variance between the data points. In the PCA plot (figure3-3B) the distance between the data points represent the degree of variance between the datasets. Separation along the two principle components that explain the majority of the variance (PC1 and PC2) result in segregation depending on the cell type. On the whole, the different samples from the same leukaemia or cell type cluster together. Along PC3, however, the two t(3;21) patients are further apart compared to the other pairs of samples. Absolute gene expression levels of genes, differentially expressed as compared to normal CD34+ PBSCs (1.5

fold), are represented in a heat map in figure 3-3C. This result demonstrates distinct gene expression patterns for each type of CBF leukaemia.

Finally, we wanted to know if the distinct DHS profiles seen in each CBF leukaemia directly result in differences in expression of the neighbouring genes. To this end we plotted the DNase-seq tag counts of t(3;21) patient 2 and t(8;21) patient 1, according to fold change in DNase-seq signal between the two patients (figure3-3D). Hence, the DHSs specific to the t(3;21) leukaemia are at the top, whilst DHSs specific to the t(8;21) leukaemia are at the bottom. Along the same genomic coordinates we plotted a heatmap based on the fold-change in gene expression between t(3;21) patient 2 and either t(8;21) patient 1 or patient 2. This analysis shows that DHSs specific to t(3;21) leukaemia are neighbouring genes which are expressed in the RUNX1-EVI-1 driven leukaemia more highly than in RUNX1-ETO leukaemia. In contrast, DHSs specific to t(8;21) leukaemia are nearest genes which are specifically expressed in t(8;21) leukaemia, as compared to the t(3;21) leukaemia.

These analyses taken together confirm that the distinct DHS profile seen in t(8;21) patients, as compared to t(3;21) patients results in distinct gene expression patterns in t(8;21) leukaemia as compared to t(3;21) cells and together with the previous analysis (figure3-2) identifies unique regulators of gene expression which define each type of CBF leukaemia.

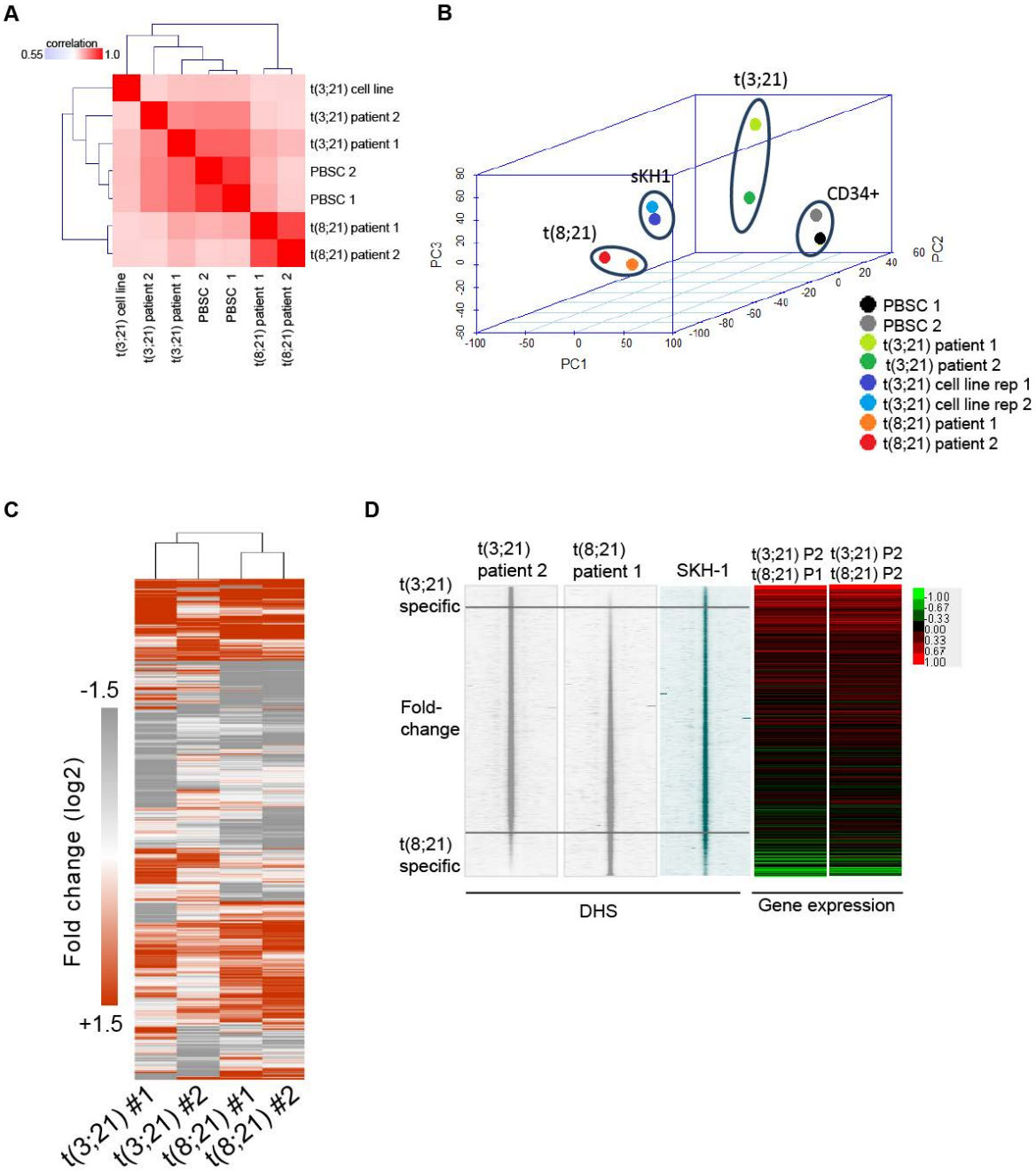


Figure 3-3 Gene expression segregates patients according to CBF translocation

A) Gene expression signals from patients with t(8;21) leukaemia cluster separately to those from patients with t(3;21) leukaemia. Correlation clustering of RNA-seq signals in t(3;21) patients and the SKH-1 cell line with t(8;21) patients and normal CD34+PBSC.

B) Principle component analysis showing separate clustering of t(3;21) patients, t(3;21) cell line, t(8;21) patients and normal CD34+PBSC based on RNA-seq signal.

C) RNA-seq in patients with t(3;21) leukaemia cluster apart from patients with t(8;21) leukaemia. Hierarchical clustering of each RNA-seq experiment by gene expression (fold change in comparison with normal CD34+ PBSCs (log2)). Clustering based on the differentially expressed genes from each patient with CBF leukaemia.

D) DHSs specific to each CBF leukaemia activate a subset of genes specific to the type of leukaemia. Ranking of DNase-seq tag counts from high to low signal within DHS from t(3;21) patient 2 and t(8;21) patient 1 (grey) and SKH-1 (turquoise). Signals were ordered in terms of t(3;21) to t(8;21) DNase-seq signal fold-change: (t(3;21) specific DHSs: top, t(8;21) specific DHSs: bottom). To the right, the heatmap shows the expression of nearest genes ordered according to the same coordinates, showing the fold change of FPKM values between t(3;21) patient 2 with either t(8;21) patient 1 or patient 2. The colour key represents gene expression fold change between patient with t(3;21) over t(8;21): (green higher in t(3;21), red higher in t(8;21). Genes more highly expressed in t(3;21) patient 2 as compared to t(8;21) patients are located at the,, genes expressed at a lower level in t(3;21) patient 2 as compared to t(8;21) patients are located at the bottom.

3.2 RUNX1 and CBF fusion proteins bind to different sites in t(3;21) and t(8;21) leukaemia

In chapter 3.1 we have demonstrated fundamental differences in the epigenetic landscape between t(3;21) and t(8;21) AML (chapter 3.1.1 and 3.1.2). These differences results in differentially expressed genes that drive the respective leukaemia phenotype. We next asked whether RUNX1 and both CBF fusion proteins bound to the same targets in the different types of AML, despite these differences in their epigenome, as both proteins retain the RUNT DNA binding domain originating from the RUNX1 portion of the translocation. We therefore performed ChIP-seq for RUNX1 and RUNX1-EVI-1 in t(3;21) SKH-1 cells to complement our pre-existing RUNX1 and RUNX1-ETO data in t(8;21) Kasumi-1 cells (Ptasinska et al., 2012). As a model of t(3;21) leukaemia we used the SKH-1 cell line, which is derived from a patient with CML in blast crisis, and

hence also the t(9;22) translocation (Mitani et al., 1994). We demonstrated in figure 3-2, that approximately 90% of the DHSs present in the SKH-1 cell line were also present in the primary patient samples with the t(3;21) translocation, validating the use of this cell line in epigenetic studies. Kasumi-1 was derived from a relapsed t(8;21) leukaemia in a paediatric patient (Asou et al., 1991), and has been extensively used as a model of t(8;21) leukaemia.

EVI-1 is not expressed in the SKH-1 cell line

In order to perform ChIP-seq of RUNX1-EVI-1 using an EVI-1 antibody, we first examined whether the untranslocated *EVI-1* was expressed in the SKH-1 cells. Figure 3-4A demonstrates that an EVI-1 antibody binds only to a protein of the size consistent with RUNX1-EVI-1 (Mitani et al., 1994) demonstrating that EVI1 is not expressed. A product of the same size was also bound by a RUNX1 antibody that detects a N-terminal epitope of RUNX1 involved in the fusion protein. This band was absent when RUNX1-EVI-1 was depleted by siRNA targeting the fusion breakpoint (see Chapter 3.4), and was present in another cell line only when it was transfected with a RUNX1-EVI-1 plasmid. We also wanted to perform ChIP-seq against the untranslocated *RUNX1* product. To do this we had to identify an antibody that bound specifically to the wild-type RUNX1 but not RUNX1-EVI-1. Figure 3-4B demonstrates that the RUNX1 C-terminal epitope antibody detects only the untranslocated *RUNX1* product, as this epitope is not present in the RUNX1-EVI-1 fusion protein.

These experiments demonstrate that an antibody for EVI-1 only binds RUNX1-EVI-1 in t(3;21) SKH-1 cells and that an antibody for the RUNX1 C-terminus only binds the product of the untranslocated *RUNX1* allele. RUNX1 and RUNX1-EVI-1 ChIP-seq in the t(3;21) SKH-1 cell line was then performed using these antibodies. High confidence ChIP-seq peaks were then identified by only selecting ChIP-seq peaks which coincide with the presence of DHSs. Figure 3-4C show that RUNX1 and both CBF fusion proteins predominantly bind to sites distal to the promoter.

Table 3-3 ChIP-seq in untreated t(3;21) SKH-1, t(8;21) Kasumi-1 and normal CD34+ PBSC
High confidence peaks are ChIP-seq peaks that overlap with a DHS.

ChIP Dataset	Total peaks (n)	High Confidence Peaks (n)
Anti-RUNX1 SKH1	15,609	10,922
Ant-EVI1 SKH1	14.992	8,947
Anti RUNX1 Kasumi1	14,972	10,080
Anti-ETO Kasumi1	14,765	9,435
Anti-RUNX1 CD34+PBSC	13,951	11,226

3.2.1 Binding sites for RUNX1 and CBF fusion proteins differ between t(8;21) and t(3;21) leukaemia

RUNX1 and CBF fusion protein targets which are shared between both t(3;21) and t(8;21) leukaemia include enhancers for *SPI1* and *CSF1R* (Follows et al., 2005, Leddin et al., 2011) (figure 3-4E). *SPI1* encodes the transcription factor PU.1 which is expressed in macrophages and B cells (Klemsz et al., 1990, DeKoter and Singh, 2000) and acts in opposition to GFI-1 to dictate differentiation of myeloid cells into either neutrophils or macrophages (Dahl et al., 2006, Dahl et al., 2003). An upstream regulatory element (URE) bound by RUNX1 and CBF fusion protein in both types of AML, has previously been shown to be important in maintaining myeloid specific expression of *SPI1* (Li et al., 2001). *CSF1R* encodes a receptor for the cytokine M-CSF and is essential for macrophage development (Dai et al., 2002). An enhancer for *CSF1R*, bound by both RUNX1 and CBF fusion proteins in both leukaemias, termed "fms intronic regulatory element" (FIRE) has previously been deleted in transgenic mouse models, whereby, the expression of *CSF1R* is reduced (Himes et al., 2001). In keeping with the similarities in both RUNX1 and CBF fusion protein binding in both leukaemias at these loci, *SPI1* expression is similar between t(3;21) and t(8;21) cells (figure3-4D), whilst *CSF1R* is not expressed in either cell type.

In contrast, both *GATA2* and *MEIS1* are more highly expressed in t(3;21) than in t(8;21) cells (figure3-4E). *GATA2* levels, as previously mentioned, are critical for correct HSC homeostasis. Similarly, whilst *Meis1* knockout mice are capable of forming HSCs, they demonstrate defects in competitively repopulating lethally irradiated mice (Hisa et al., 2004). In keeping with the differences in DHS profile at these loci (for *MEIS1* figure 3-1A), between t(8;21) and t(3;21) AML patients. Both RUNX1 and RUNX1-EVI-1 are bound at the promoter at *MEIS1* whilst in t(8;21) cells RUNX1 and RUNX1-ETO do bind there (figure 3-4F). At *GATA2,* RUNX1 is bound to the proximal promoter in t(3;21) but not in t(8;21) cells (Rodrigues et al., 2012). RUNX1 and RUNX1-ETO bind a distal enhancer shown to regulate *GATA2* (Groschel et al., 2014) but this is not bound by neither RUNX1 nor RUNX1-EVI-1 in the t(3;21) cells.

The differences in RUNX1-ETO and RUNX1-EVI-1 binding, alongside differences in RUNX1 binding in each leukaemia are therefore associated with differential expression of key haematopoietic transcription factors. This feature may be responsible for the differences in phenotype of the leukaemia seen in the patients (see chapter 1.7.1). Furthermore, differences in RUNX1 and CBF fusion protein binding between each leukaemia appears to be associated with differences in the DHSs between t(3;21) and t(8;21) AML.

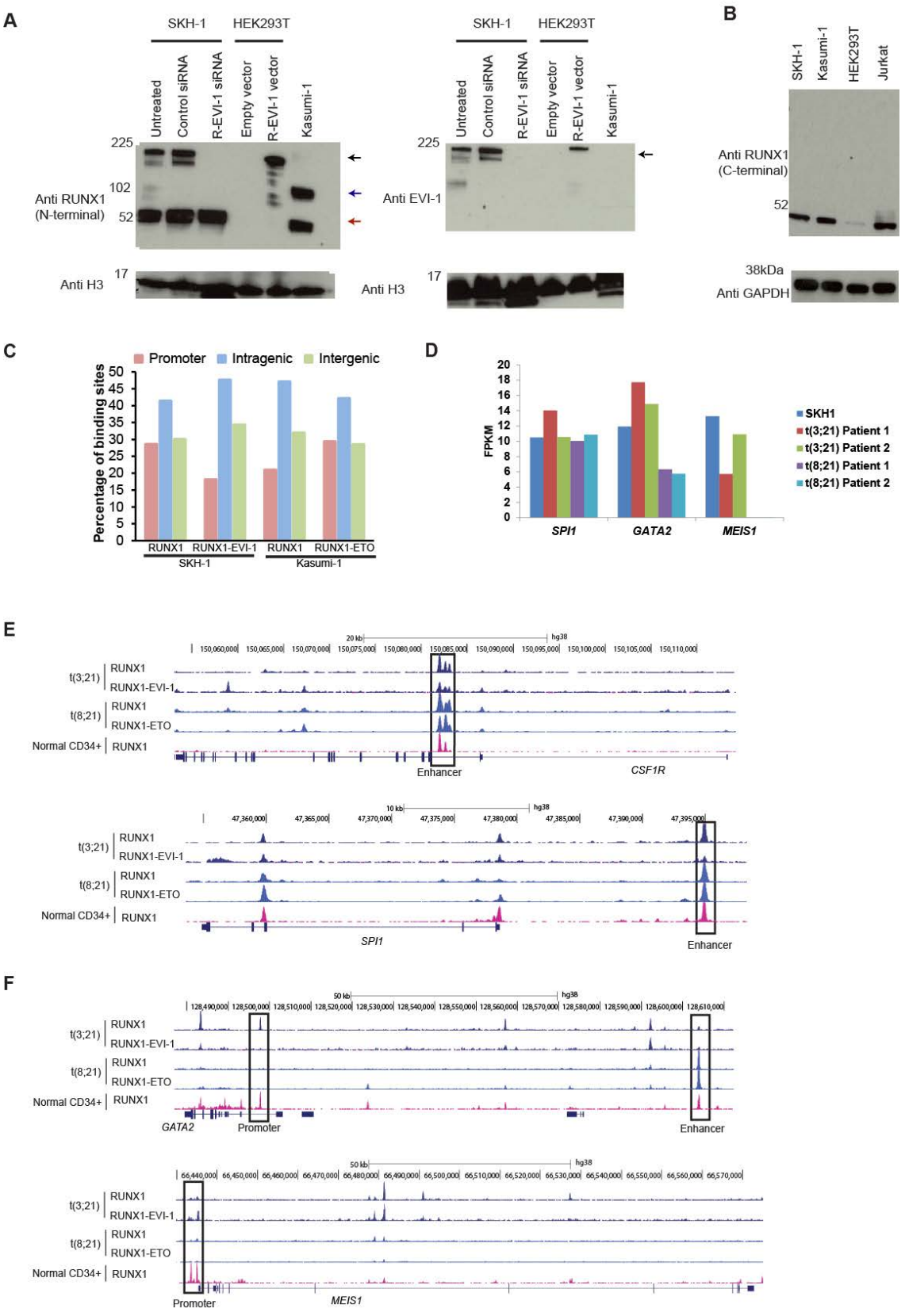

A
SKH-1
HEK293T
Untreated
Control siRNA
R-EVI-1 siRNA
Empty vector
R-EVI-1 vector
Kasumi-1
225
102
52
Anti RUNX1
(N-terminal)
Anti H3
17
Anti EVI-1
B
SKH-1
Kasumi-1
HEK293T
Jurkat
Anti RUNX1
(C-terminal)
52
38kDa
Anti GAPDH
C
Promoter
Intragenic
Intergenic
Percentage of binding sites
RUNX1
RUNX1-EVI-1
RUNX1-ETO
SKH-1
Kasumi-1
D
FPKM
SKH1
t(3;21) Patient 1
t(3;21) Patient 2
t(8;21) Patient 1
t(8;21) Patient 2
SPI1
GATA2
MEIS1
E
t(3;21)
t(8;21)
Normal CD34+
Enhancer
CSF1R
SPI1
F
GATA2
Promoter
MEIS1

Figure 3-4 Common and distinct gene targets of RUNX1 and CBF fusion proteins in t(3;21) as compared to t(8;21) leukaemia
A-B) Anti EVI-1 antibody binds specifically to RUNX1-EVI-1 and anti RUNX1 (C-terminal epitope) antibody binds specifically to wild-type RUNX1.

A) Western blot with nuclear extracts. From left to right: t(3;21) SKH-1 cells (untreated or transfected with control siRNA or RUNX1-EVI-1 siRNA), HEK293T cells (transfected with either empty vector or RUNX1-EVI-1 vector), and t(8;21) Kasumi-1 cells. Size in kDa to the left of blot. Western blots were probed with either an anti-EVI-1 or an anti-RUNX1 (N-terminal epitope antibody) as indicated. Anti-EVI-1 and anti-RUNX1 (N-terminal epitope) antibodies detect a RUNX1-EVI-1 specific band (black arrow). This band is absent in SKH-1 transfected with a RUNX1-EVI-1 specific siRNA and is present in HEK293T cells only when transfected with a RUNX1-EVI-1 plasmid. In Kasumi-1 nuclear extract, EVI-1 antibody does not detect anything, whilst the N-terminal RUNX1 antibody also detects RUNX1-ETO (which is smaller than RUNX1-EVI-1) (blue arrow). The 52kDa band is wild type RUNX1 (red arrow). Anti-H3 antibody used as a loading control.

B) Western blot of whole cell lysates from SKH-1, Kasumi-1, HEK293T and Jurkat cells. Blot probed with an antibody against the c-terminal epitope of RUNX1. RUNX1 (c-terminal epitope) antibody only detects untranslocated *RUNX1* product in t(3;21) SKH-1 and t(8;21) Kasumi-1. HEK293T as negative control and Jurkat as positive control. An anti-GAPDH antibody was used to highlight the loading control.

C) Percentage of binding sites by promoter, intragenic and intergenic region for RUNX1 and EVI-1 (RUNX1-EVI-1) ChIP seq in SKH-1 and RUNX1 and ETO (RUNX1-ETO) ChIP-seq in Kasumi-1.

D) Expression of *SPI1*¸ *GATA2* and *MEIS1* by RNA-seq (FPKM values) in CD34+ purified cells from two patients with t(3;21) AML and two patients with t(8;21) AML. Alongside this are FPKM values from RNA-seq of SKH-1, (average of two technical replicates shown) for the same genes. Expression of *SPI1* is the same in both t(3;21) and t(8;21) cells. *MEIS1* is expressed in t(3;21) but not in t(8;21) cells. *GATA2* is more highly expressed in t(3;21) cells as compared to t(8;21) cells.

E-F) UCSC browser screen shots of aligned reads from ChIP-seq of RUNX1 and RUNX1-EVI-1 binding in SKH-1 cells, RUNX1 and RUNX1-ETO binding from Kasumi-1 cells and RUNX1 from normal CD34+ PBSC.

E) RUNX1 binding in both t(3;21) and t(8;21) cells, and both RUNX1-EVI-1 and RUNX1-ETO bind at enhancers in *CSF1R*, and *SPI1* loci.

F) RUNX1 and RUNX1-EVI-1 bind at *GATA2* and *MEIS1* loci in t(3;21) but these genes are not bound by RUNX1 or RUNX1-ETO in t(8;21) cells. RUNX1 and RUNX1-ETO bind at a distal enhancer for *GATA2*.

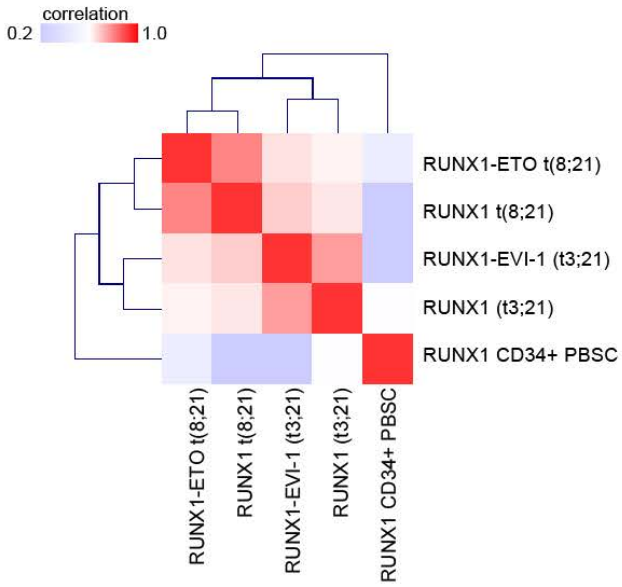
3.2.2 Binding sites differ because unique motifs are found in specific RUNX1, RUNX1-EVI-1 and RUNX1-ETO targets

The above-described experiments show that despite both CBF fusion proteins retaining the RUNT DNA binding domain they bound to distinct binding sites for each fusion protein. To quantify these differences of the overall binding patterns of RUNX1 and each CBF fusion protein on a genome wide level, we performed a correlation analysis of the peaks of all four ChIP-seq experiments (RUNX1 and RUNX1-ETO in t(8;21) Kasumi-1 cells and RUNX1 and RUNX1-EVI-1 in t(3;21) SKH-1 cells).

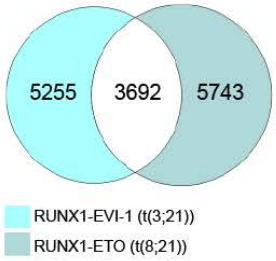
Figure 3-5A shows that RUNX1 binding sequences cluster according to their cell line of origin and with the binding sites of the respective fusion proteins, which, in turn, cluster independently for t(8;21) cells and t(3;21) cells (figure 3-5A). In addition, we compared the RUNX1 binding pattern with that of normal PBSCs (Cauchy et al., 2015). Again, RUNX1 ChIP-seq binding sites from normal PBSCs cluster independently to either leukaemia. This suggests that the RUNX1 binding pattern was unique to either t(3;21) or t(8;21) leukaemia. Indeed, RUNX1 binding pattern bore a closer similarity to the CBF fusion protein within that cell line, than to RUNX1 in the other CBF leukaemia.

To understand the mechanisms driving these differences in binding between the CBF fusion proteins in each type of AML we identified the sites bound specifically by each fusion protein and performed a de novo motif analysis at these sites using HOMER (Heinz et al., 2010). This analysis revealed that GATA motifs were enriched in RUNX1-EVI-1 specific binding sites, whilst CEBP and E-box motifs were enriched in RUNX1-ETO specific binding sites (figure 3-5B). This result was consistent with our de novo motif analyses in the DHSs specific to each leukaemia (figure 3-2). We also compared the motifs enriched in the binding sites for RUNX1 and each CBF fusion protein (figure 3-5C and D) to determine whether or not a particular motif characterises specific CBF fusion protein binding sites in each leukaemia. Again, there was a large percentage of E-box motifs in the sites specific to RUNX1-ETO as compared to RUNX1 in t(8;21) cells.

A



B



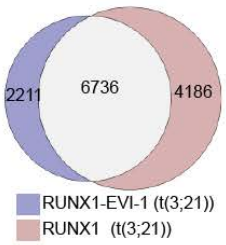
RUNX1-EVI-1 unique distal peaks

motif	Match	Score	%target
	ETS	1e-575	51.7
	AP1	1e-171	15.7
	RUNX1	1e-160	38.3
	GATA	1e-112	17.9

RUNX1-ETO unique distal peaks

motif	Match	Score	%target
	ETS	1e-358	51.2
	RUNX1	1e-320	51.2
	C/EBP	1e-91	21.7
	AP1	1e-85	15.1
	E-box	1e-58	35.2

C



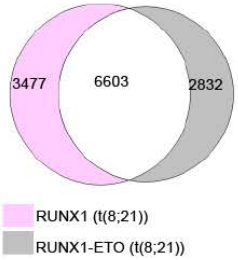
RUNX1-EVI-1 unique distal peaks

motif	Match	Score	%target
	ETS	1e-194	47.0
	AP1	1e-72	17.3
	CTCF	1e-56	10.6
	RUNX	1e-50	28.3

RUNX1 unique distal peaks

motif	Match	Score	%target
	ETS	1e-218	48.2
	AP1	1e-97	18.6
	RFX1	1e-91	17.2
	RUNX1	1e-50	28.8

D



RUNX1-ETO unique distal peaks

motif	Match	Score	%target
	ETS	1e-107	37.2
	AP1	1e-66	15.2
	RUNX1	1e-41	49.6
	Ebox	1e-27	38.4

RUNX1 unique distal peaks

motif	Match	Score	%target
	RUNX	1e-307	44.8
	ETS	1e-284	34.6
	CTCF	1e-77	7.8
	AP1	1e-46	22.2

Figure 3-5 RUNX1 and CBF fusion proteins binding partially overlaps within each type of AML and binding sites contain transcription factor binding motifs unique to each CBF fusion protein.
A) RUNX1 and RUNX1-ETO binding sites in t(8;21) leukaemia cluster separately to RUNX1 and RUNX1-EVI-1 in t(3;21) leukaemia. Hierarchical clustering of correlation between ChIP-seq experiments: RUNX1 and RUNX1-ETO in t(8;21) Kasumi-1, RUNX1 and RUNX1-EVI-1 in t(3;21) SKH-1 and RUNX1 in CD34+ PBSC.

B-D) Venn diagram of overlaps between ChIP-seq peaks from different experiments with tables comparing de novo motif analyses of distal sites unique to each transcription factor:

B) Comparison of RUNX1-EVI-1 binding sites in t(3;21) SKH-1 with RUNX1-ETO binding sites in t(8;21) Kasumi-1 cells (Ptasinska et al., 2012). The motif analysis on the right shows GATA motifs in RUNX1-EVI-1 unique binding sites (left) and CEBP and E-Box motifs in RUNX1-ETO unique binding sites.

C) Comparison of RUNX1-EVI-1 binding sites with RUNX1 sites in t(3;21) SKH-1. The motif analysis on the right shows CTCF in RUNX1-EVI-1 unique binding sites and RFX1 in RUNX1 unique binding sites.

D) Comparison of RUNX1 1 binding sites with RUNX1-ETO in t(8;21) Kasumi-1 cells. The motif analysis on the right shows E-Box in RUNX1-ETO unique binding sites and CTCF in RUNX1 unique binding sites.

The differences between RUNX1-EVI-1 and RUNX1-ETO targets could be confirmed by looking at the percentage overlap of ChIP-seq peaks. Figure 3-6A shows that for each leukaemia type, a large percentage of binding sites are shared between RUNX1 and the different CBF fusion proteins. In contrast, even for the same transcription factor (RUNX1) the percentage of shared targets in t(3;21) and t(8;21) leukaemia is comparatively reduced.

In order to understand the reasons for this pattern of binding site overlap we examined whether particular motifs were enriched in binding sites specific to each experiment. This analysis involved the identification of peaks specific for each ChIP-seq experiment in a pairwise comparison of all experiments. We then identified known transcription factor binding motifs in these specific peaks. In order to determine whether or not motifs were truly enriched in one particular ChIP seq experiment, they were then compared to the union of all the discovered motifs in a pairwise comparison (analysis scheme shown in figure 3-

6B). This analysis was then subjected to a clustering analysis, with a resultant heatmap showing which known motifs were enriched as compared to their background occurrence (figure 3-6C). This analysis highlighted transcription factor families shaping chromatin accessibility specific for each particular cell type. The figure shows that enriched motifs in t(8;21) clustered separately to enriched motifs in t(3;21) leukaemia. Known motifs enriched in RUNX1 peaks in t(8;21) leukaemia include RUNX, GFI1B and CEBP. In contrast, motifs enriched in RUNX1 or RUNX1-EVI-1 peaks in t(3;21) leukaemia included GATA, HOXA9 and AP1. Enrichment for E-box motifs in RUNX1-ETO specific peaks was specific to this fusion protein and was not found in any of the other ChIP-seq peaks.

The separate segregation of enriched motifs in transcription factor binding sites between t(3;21) and t(8;21) leukaemia in this analysis, is consistent with the separated clustering in the correlation clustering of the ChIP-seq experiments in figure 3-5A and 3-6A. Furthermore, the enriched motifs found in the simple pairwise comparisons between ChIP-seq experiments in figure 3-5B and D are replicated as part of the results of this analysis in figure 3-6C. Together these analyses suggest that interactions of RUNX1-EVI-1 and RUNX1-ETO with different transcription factors in each leukaemia results in differences in binding pattern of each CBF fusion protein.

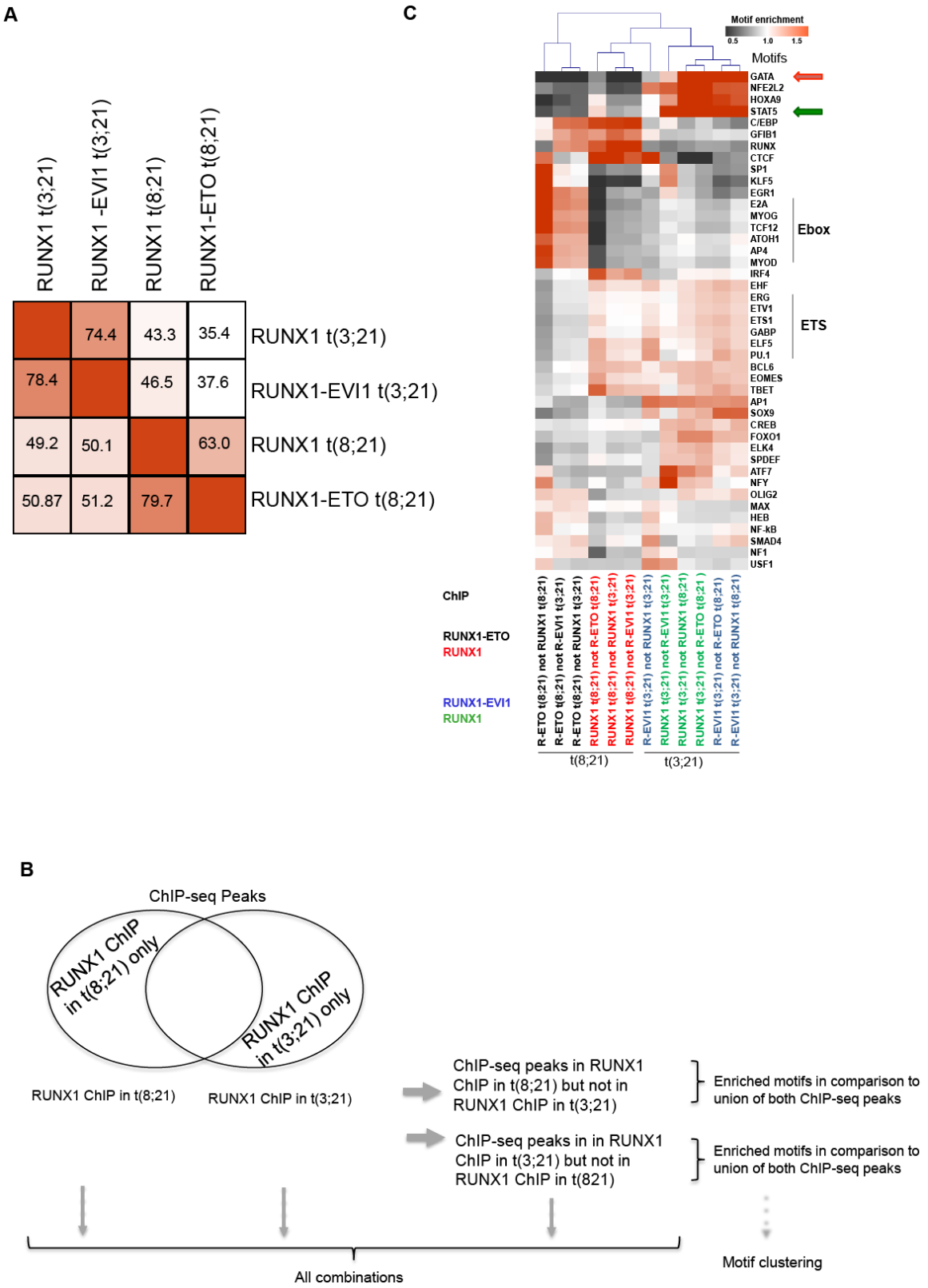

A
RUNX1 t(3;21)
RUNX1 -EVI1 t(3;21)
RUNX1 t(8;21)
RUNX1-ETO t(8;21)
74.4
43.3
35.4
RUNX1 t(3;21)
78.4
46.5
37.6
RUNX1-EVI1 t(3;21)
49.2
50.1
63.0
RUNX1 t(8;21)
50.87
51.2
79.7
RUNX1-ETO t(8;21)
C
Motif enrichment
0.5
1.0
1.5
Motifs
GATA
NFE2L2
HOXA9
STAT5
C/EBP
GFIB1
RUNX
CTCF
SP1
KLF5
EGR1
E2A
MYOG
TCF12
ATOH1
AP4
MYOD
Ebox
IRF4
EHF
ERG
ETV1
ETS1
GABP
ELF5
PU.1
BCL6
EOMES
TBET
AP1
SOX9
CREB
FOXO1
ELK4
SPDEF
ATF7
NFY
OLIG2
MAX
HEB
NF-kB
SMAD4
NF1
USF1
ETS
ChIP
RUNX1-ETO
RUNX1
RUNX1-EVI1
RUNX1
R-ETO t(8;21) not RUNX1 t(8;21)
R-ETO t(8;21) not R-EVI1 t(3;21)
R-ETO t(8;21) not RUNX1 t(3;21)
RUNX1 t(8;21) not R-ETO t(8;21)
RUNX1 t(8;21) not RUNX1 t(3;21)
RUNX1 t(8;21) not R-EVI1 t(3;21)
R-EVI1 t(3;21) not RUNX1 t(3;21)
RUNX1 t(3;21) not R-EVI1 t(3;21)
RUNX1 t(3;21) not RUNX1 t(8;21)
RUNX1 t(3;21) not R-ETO t(8;21)
R-EVI1 t(3;21) not R-ETO t(8;21)
R-EVI1 t(3;21) not RUNX1 t(8;21)
t(8;21)
t(3;21)
B
ChIP-seq Peaks
RUNX1 ChIP in t(8;21) only
RUNX1 ChIP in t(3;21) only
RUNX1 ChIP in t(8;21)
RUNX1 ChIP in t(3;21)
ChIP-seq peaks in RUNX1 ChIP in t(8;21) but not in RUNX1 ChIP in t(3;21)
Enriched motifs in comparison to union of both ChIP-seq peaks
ChIP-seq peaks in in RUNX1 ChIP in t(3;21) but not in RUNX1 ChIP in t(821)
Enriched motifs in comparison to union of both ChIP-seq peaks
All combinations
Motif clustering

Figure 3-6 Clustering identifies unique patterns of enriched transcription factor binding motifs within RUNX1 and CBF fusion protein binding sites within each type of AML

A) Overlap between RUNX1 and core-binding factor fusion protein within, but not between, each leukaemia. Percentage of overlapping peaks at distal sites: pair wise comparison of wild-type RUNX1, RUNX1-ETO and RUNX1-EVI-1 in t(3;21) and t(8;21) leukaemia.

B) Analysis scheme for pairwise analysis of ChIP-seq data for each CBF leukaemia. Peak overlap was performed for each combination of ChIP-seq peak data. This identified peaks unique to each type of AML. Enrichment score was calculated by the level of motif enrichment in the unique peaks, as compared to union of peaks in the pair of experiments. Clustering analysis performed on the motif enrichment.

C) Hierarchical clustering of motifs discovered in the pairwise comparison between RUNX1 and CBF fusion ChIP-seq peaks between t(3;21) and t(8;21) leukaemia. The heat map representing motif enrichment as described in B). Unique peaks from ChIP-seq experiments in t(8;21) cells cluster separately to unique peaks from ChIP-seq experiments in t(3;21) cells which identifies enriched motifs unique to each ChIP-seq experiment. For example, enrichment of E-Box motifs in RUNX1-ETO ChIP-seq in comparison to RUNX1 ChIP-seq in t(8;21) cells (marked on diagram). GATA and STAT5 motifs (highlighted) enriched in RUNX1-EVI-1 and RUNX1 binding sites in t(3;21) leukaemia. ETS motif (highlighted) enriched in RUNX1 binding sites in t(8;21) cells and in RUNX1 and RUNX1-EVI-1 binding sites in t(3;21) cells. Bound RUNX1 and CBF fusion protein complexes differ between primary t(3;21) and t(8;21) cells

The differences in motif enrichment between the different CBF fusion protein and RUNX1 binding sites suggest that they may bind cooperatively with differentially expressed transcription factors in each leukaemia (figure 3-6C). However, this analysis does not prove that the motifs are actually occupied by a transcription factor, nor does it necessarily demonstrate that they are relevant in patient samples, and not only in the cell line in which the ChIP-seq experiments were performed. Therefore, we analysed high read depth DNase-seq data from t(3;21) and t(8;21) AML patient samples with our Wellington footprinting algorithm (Piper et al., 2013) to determine whether or not these motifs were directly bound by transcription factors in the patient samples, at RUNX1 or CBF fusion protein bound sites revealed by ChIP-seq in the cell lines. Wellington

identifies DNA-Sequences protected from DNaseI digestion within DNaseI hypersensitive sites.

To identify co-localising transcription factors we calculated the distance between transcription factor occupied ("footprinted") motifs and motifs for defined transcription factors as co-localising if they were within 50bp of each other. We also generated a pooled set of all footprinted motifs, identified from t(3;21), t(8;21) and CD34+ cells (Ptasinska et al., 2014, Cauchy et al., 2015). We then performed a bootstrapping analysis to assess whether these footprinted motifs co-localise more than would be expected from their occurrence in this pooled set of all occupied motifs. Bootstrapping analysis is based on re-sampling the pool of all occupied motifs in order to determine probability of the motifs co-localising simply by chance. Finally after hierarchical clustering of the footprint co-occurrence, the z-scores were displayed as a heatmap (figure 3-7A-D). The red intensity correlates with increasing likelihood that the two transcription factors binding co-localise.

Figure 3-7A and B shows the results of this analysis for RUNX1 bound sites in cells of patients with a t(3;21) and t(8;21) leukaemia, respectively. In these analyses, co-localising footprinted motifs in the patients were identified at RUNX1 binding sites identified in the cell lines using ChIP-seq. The analysis shows that at RUNX1 bound sites in t(3;21) patient samples, occupied ERG, PU.1, GATA and AP-1 motifs co-localise. In contrast, in t(8;21) patients occupied RUNX1 motifs co-localise with ERG and CEBP transcription factors.

Figure 3-7C and D show the same analysis for RUNX1-EVI-1 and RUNX1-ETO in t(3;21) and t(8;21) cells, respectively. This analysis shows that in t(8;21) patient cells, occupied E-Box sites co-localise with RUNX and ETS (ERG/PU.1) bound transcription factors at RUNX1-ETO binding sites. This is in keeping with previous mass-spectrometry based data demonstrating that RUNX1-ETO forms a complex with E-box proteins such as HEB and E2A (Sun et al., 2013). However, it was not known which transcription factors co-localise with RUNX1-EVI-1and to facilitate its binding at target genes. Here we show that, in t(3;21) patient cells, at RUNX1-EVI-1 bound sites ETS (ERG/PU.1) factor motifs are

clustering together with AP-1 and GATA motifs (figure 3-7C). Interestingly, RUNX motifs were not part of this cluster.

Given the high enrichment of the GATA motif in both RUNX1 and RUNX1-EVI-1 binding sites in t(3;21) cells, but not in RUNX1 and RUNX1-ETO in t(8;21) cells, we queried which GATA factor was likely to bind there. Although RUNX1-EVI-1 can bind GATA motifs in vitro (Delwel et al., 1993), the consensus motif they identified is a longer composite motif that would not explain the presence of the GATA motifs in RUNX1 bound regions. Of the GATA family of transcription factors, in t(3;21) cells only GATA1 and 2 are expressed, whereby GATA2 is more highly expressed than GATA1 (figure 3-7E), and is more highly expressed in t(3;21) cells than in t(8;21) cells. To test whether our identified DHS are generally capable of binding GATA2, we analysed a GATA2 ChIP-seq data set from another leukaemic cell-line TF1 (Mazumdar et al., 2015) and filtered GATA2 binding sites against DHSs in the SKH-1 cell line. Following this we show a substantial overlap of these GATA2 binding sites with those of RUNX1 and RUNX1-EVI-1 (figure 3-7F).

Taken together, these data suggests that RUNX1, RUNX1-EVI-1 and RUNX1-ETO occupy different targets in each CBF leukaemia and form different transcription factor complexes in both cell lines and purified primary leukaemic blasts from patients. This finding provides a mechanistic explanation to the difference between each CBF leukaemia.

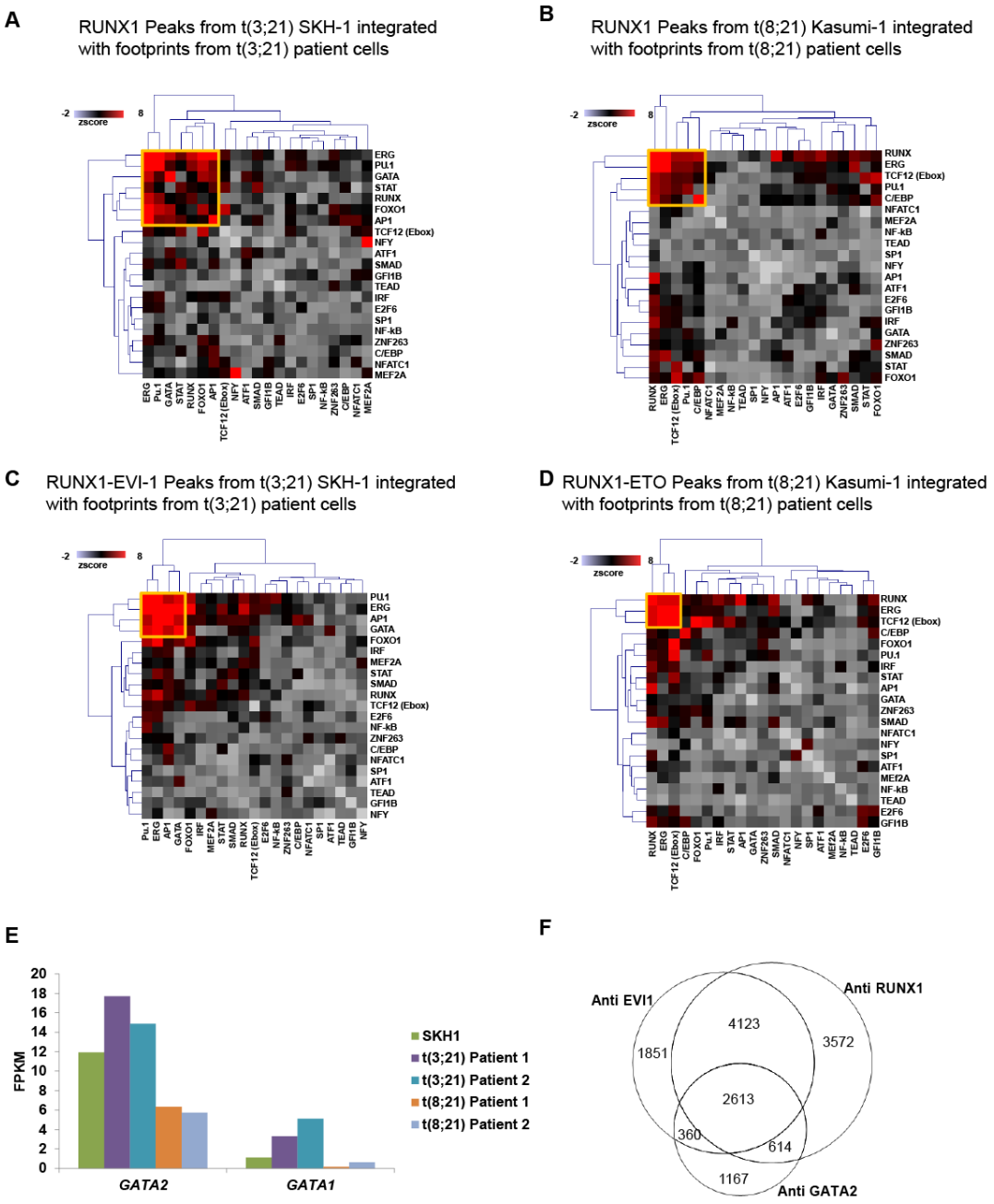


Figure 3-7 Identification of different RUNX1 and CBF fusion protein complexes in t(8;21) compared to t(3;21) patient cells

A-B) Heatmap showing co-localisation of occupied transcription factor binding motifs in t(3;21) or t(8;21) cells in association with RUNX1 binding in each leukaemia. Enrichment (red) of occupied GATA, AP-1, ERG and PU.1 motifs at RUNX1 bound sites in t(3;21) patient samples. In contrast, enrichment (red) of RUNX, ERG and CEBP footprinted motifs at RUNX1 bound sites in t(8;21) patient samples. Bootstrap analysis of footprinted motifs in RUNX1 binding sites from t(3;21) or t(8;21) leukaemia. Footprint probabilities from either t(3;21)

patient 2 or t(8;21) patient 1. RUNX1 binding sites from ChIP-seq in A) t(3;21) SKH-1 and B) t(8;21) Kasumi-1 cell line.

C-D) Heatmap showing co-localisation of transcription factor binding motifs in t(3;21) or t(8;21) cells in association with CBF fusion protein binding in each leukaemia. Enrichment (red) of occupied GATA, AP-1, ERG and PU.1 motifs at RUNX1-EVI-1 bound sites in t(3;21) patient samples in contrast to the enrichment (red) of footprinted RUNX, ERG and E-Box motifs at RUNX1-ETO bound sites in t(8;21) patient samples. Bootstrap analysis of footprinted motifs from patients within RUNX1-EVI-1 or RUNX1-ETO bound sites. C) RUNX1-EVI-1 binding sites from t(3;21) SKH-1 cell line and D) RUNX1-ETO binding sites from t(8;21) Kasumi-1 cell line. Footprint probabilities from either t(3;21) leukaemia patient 2 (for RUNX1-EVI-1 targets) or t(8;21) leukaemia patient 1 (for RUNX1-ETO targets).

E) Expression of GATA2 and GATA1 in SKH1 (average of independent replicates), t(3;21) and t(8;21) patients (two patients for each CBF leukaemia) based on FPKM values from RNA-seq.

F) Peak overlaps from anti-GATA2 ChIP-seq in TF1 cells (previously published) (Mazumdar et al., 2015) and from SKH-1 cells, anti-EVI-1 (RUNX1-EVI-1) and anti-RUNX1 ChIP-seq (this project). GATA2 binding sites from TF1 cells co-localise with DHSs in SKH-1 cells. Substantial overlap of binding sites for all three transcription factors.

3.3 Characterization of RUNX1-EVI-1 and RUNX1-ETO mediated gene deregulation in normal human precursor cells

Apart from the difference in fusion partners between RUNX1-EVI-1 and RUNX1-ETO, the two types of AML differ in terms of cooperating mutations. The differences in cooperative mutations between the t(3;21) and t(8;21) primary AML samples are listed in table 2-1. The t(3;21) SKH-1 cell line also has a t(9;22) which produces the fusion kinase BCR-ABL (Mitani et al., 1994). Similarly, the t(8;21) Kasumi-1 cell line carries mutations that also affect tyrosine kinase signalling, spliceosome and polycomb regulation (KIT N822K, SRSF12 R179Q, ASXL1 G646fs*12) (COSMIC catalogue of mutations, https://cansar.icr.ac.uk/cansar/cell-lines/KASUMI-1/mutations/ accessed 18th July 2016). We therefore wanted to determine the extent to which the differences in the epigenome and binding of RUNX1 and the CBF fusion proteins were a result of these cooperative mutations, or which were a direct

consequence of the different structures of CBF fusion proteins themselves. To address this question we used lentiviral transduction to express either RUNX1-ETO or RUNX1-EVI-1 in CD34+ purified PBSCs from healthy donors, and aimed to profile their chromatin accessibility using ATAC-seq (assay for transposase accessible chromatin using sequencing (Buenrostro et al., 2013)) and measure gene expression by RNA-seq.

Using amplicons specific to either RUNX1-EVI-1 or RUNX1-ETO, we show by RT-qPCR, that we can successfully express each CBF fusion gene in CD34+PBSCs (figure 3-8A). Figure 3-7B shows flow cytometry of CD34+PBSCS expressing GFP from either an empty, RUNX1-ETO or RUNX1-EVI-1 vector (figure 3-8B).

Successfully transduced RUNX1-EVI-1 CD34+ PBSCs could not persist in long-term liquid culture (figure 3-8C), in contrast to RUNX1-ETO transduced CD34+ PBSCs. A caveat of the experiments in figure 3-8C-E was that MOI of RUNX1-EVI-1 virus titre tended to be far lower than either empty or RUNX1-ETO virus titres due to the size of the RUNX1-EVI-1 transgene (RUNX1-EVI-1 vector: 14425bp, RUNX1-ETO vector: 12557bp and empty vector: 10209bp). Currently we therefore cannot say whether the inability of outgrowth of transduced clones is a result of inefficient transduction or toxicity of the RUX1-EVI1 fusion protein in a normal cellular background.

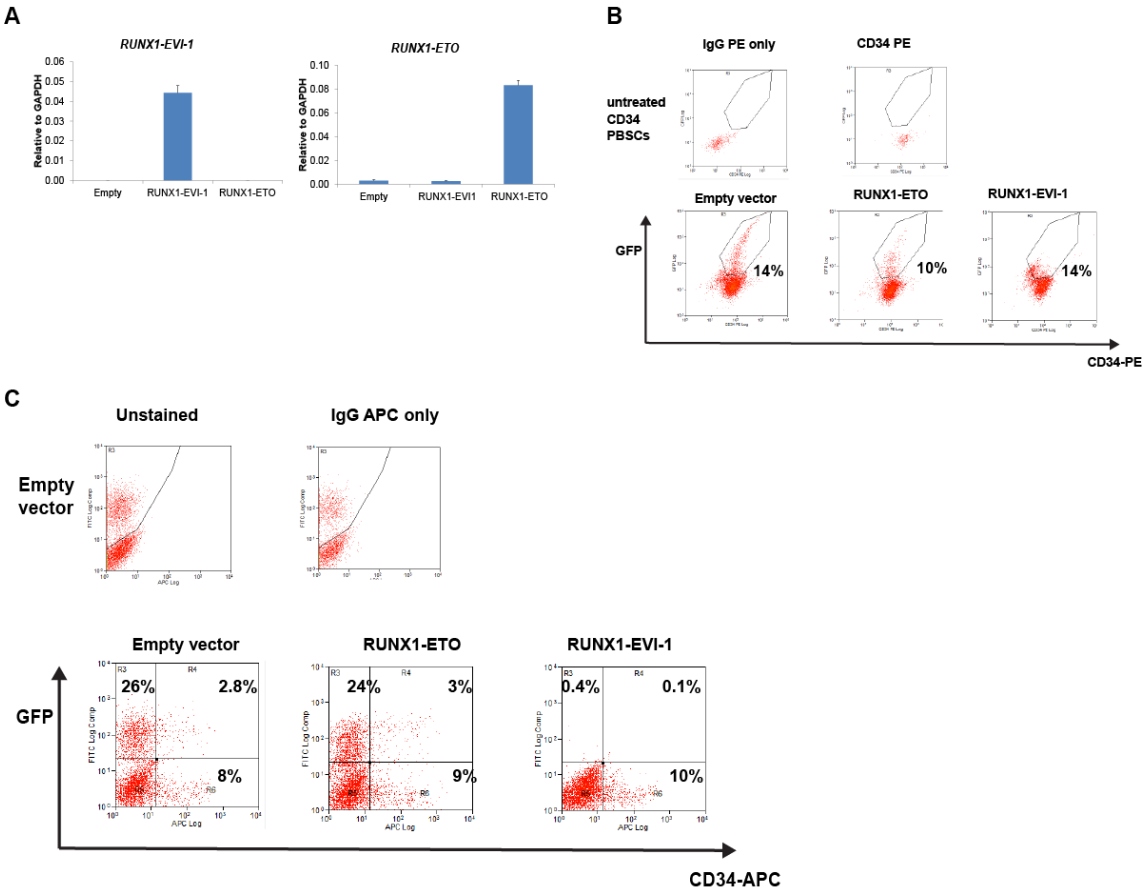


Figure 3-8 RUNX1-ETO but not RUNX1-EVI-1 expressing normal CD34+ PBSCs persist in long term culture

A)-B) RUNX1-ETO and RUNX1-EVI-1 can be expressed in purified CD34+ PBSC by lentiviral transduction. CD34+ PBSCs transduced with empty vector, or vector with either RUNX1-EVI-1 or RUNX1-ETO transgene. Successfully transduced cells were marked by expression of GFP from the lentiviral backbone.

A) RT-qPCR for either RUNX1-ETO or RUNX1-EVI-1 in CD34+ PBSCs transduced with either empty, RUNX1-ETO or RUNX1-EVI-1 pSIEW vector. Expression levels were calculated relative to *GAPDH*.

B) Flow cytometry of CD34+ PBSCs 2 days after transduction with either empty, RUNX1-ETO or RUNX1-EVI-1 pSIEW vector, stained with CD34-PE. Also shown are untreated CD34+ PBSCs either unstained or stained with IgG PE.

C) RUNX1-ETO, but not RUNX1-EVI-1 persist in PBSC in long term culture. Flow cytometry of CD34+ PBSCs 29 days after transduction with either empty, RUNX1-ETO or RUNX1-EVI-1 pSIEW vector, stained with CD34-PE.

3.4 RUNX1-EVI-1 is necessary to maintain t(3;21) leukaemia

Our inability to maintain RUNX1-EVI-1 expressing CD34+ PBSCs in contrast to RUNX1-ETO expressing CD34+ PBSCs may be consistent with the observation that RUNX1-EVI-1 is often seen in leukaemias as a second mutation (section 1.73), and therefore requires the cooperation with other mutations. It is known that RUNX1-ETO is a mutation that precedes the acquisition of others (Wiemels et al., 2002, Miyamoto et al., 2000), whilst RUNX1-EVI-1 is acquired in the presence of other preceding mutations (Nukina et al., 2014, Paquette et al., 2011, Rubin et al., 1990, Rubin et al., 1987). Given these observation, we wished to determine whether RUNX1-EVI-1 was required to maintain the full leukaemic potential of t(3;21) cells. To this end, we designed siRNA that targeted the fusion junction between RUNX1 and EVI-1, thereby specifically depleting RUNX1-EVI-1 but leaving the product of the untranslocated *RUNX1* allele intact (figure 3-9A). The decrease in protein levels after RUNX1-EVI-1 siRNA transfection is shown in figure 3-4A, whilst the decrease in mRNA levels is shown in figure 3-9A. We then repeated this transfection to maintain the knockdown of RUNX1-EVI-1 (figure 3-9B).

RUNX1-EVI1 knock-down led to a concomitant decrease in the stem cell and early progenitor marker CD34 (figure 3-9C-D) which was associated with down regulation of genes important for stem cell function and associated with malignant transformation: *HOXA9, MEIS1* and *GATA2* (Kroon et al., 1998, Schnabel et al., 2000, Lawrence et al., 2005, Rodrigues et al., 2005) (as mentioned in chapter 3.1.1). In contrast, genes expressed in differentiated myeloid cells increased in levels, and continued to do so at the last recorded measurement of the time course at day 14: *CSF1R*, *CTSG*, *MPO* (Borregaard et al., 2007, Dai et al., 2002) (figure 3-9E-F). Despite these increases in expression of genes required for monocyte function, the CD14 surface marker was not upregulated following RUNX1-EVI-1 knockdown (figure3-9G).

Thus, although RUNX1-EVI-1 may not be a primary mutation in the development of AML, it is required to maintain the early block in differentiation which is a hallmark of this disease. This would be consistent with other well-characterised mutations in AML, such as *NPM1* (Falini et al., 2005), which are secondary mutations but play a critical role in the development of the disease (Papaemmanuil et al., 2016, Mupo et al., 2013).

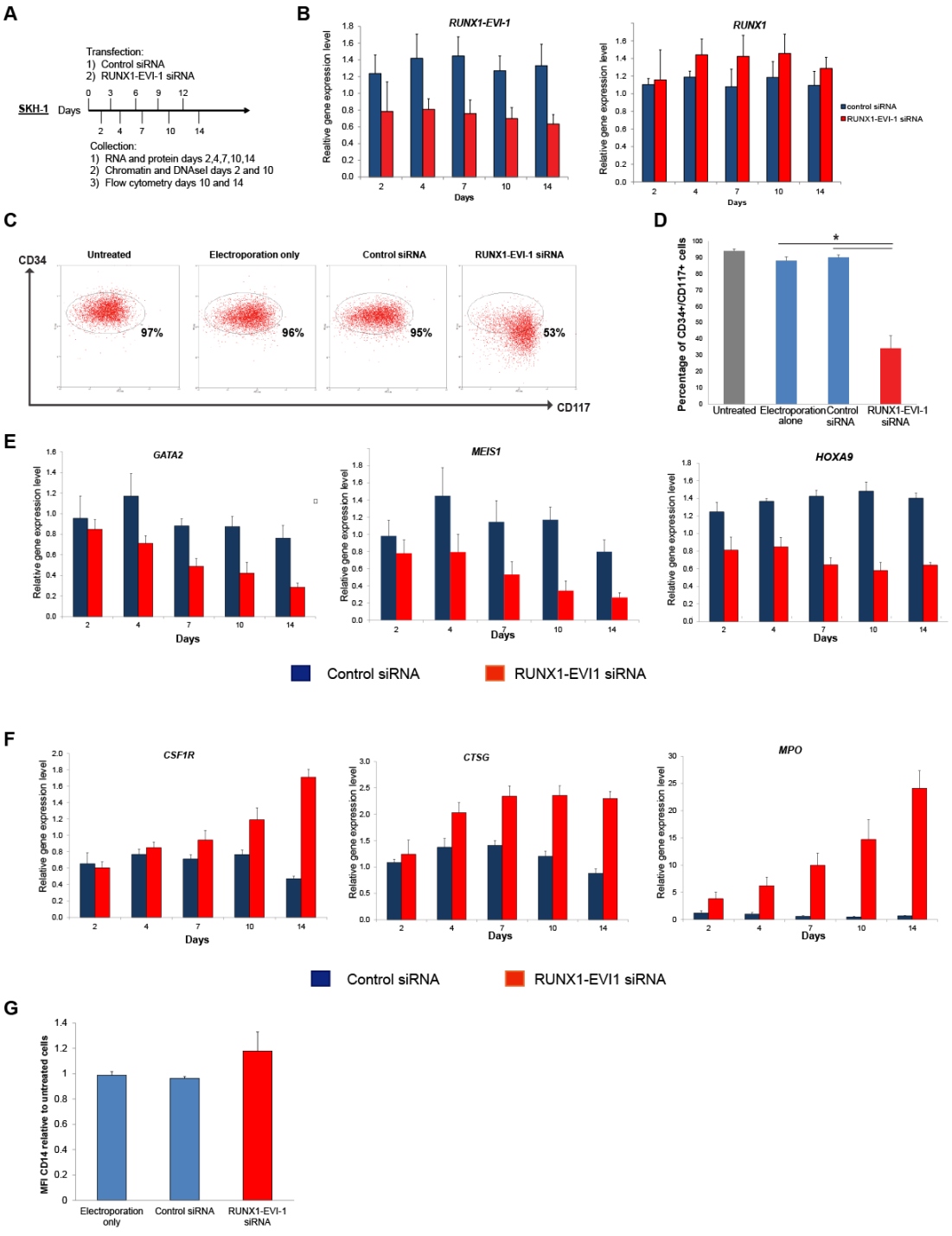


Figure 3-9 siRNA depletion of RUNX1-EVI-1 results in differentiation of t(3;21) SKH-1 cells

A) Experimental scheme. Cells were transfected with siRNA every 72 hrs and cells were collected for RNA on days 2, 4, 7, 10, 14. DNaseI treated cells and chromatin were collected on days 2 and 10. Flow cytometry was performed on day 10 and 14.

B) RUNX1-EVI-1 mRNA decrease in SKH-1 cells treated with RUNX1-EVI-1 siRNA, as compared to control siRNA treatment. RUNX1 mRNA levels are unaffected by RUNX1-EVI-1 siRNA as compared to control siRNA. RT-qPCR showing mRNA levels of the indicated genes relative to *GAPDH* and normalised to untreated cells. *RUNX1-EVI-1* or *RUNX1* mRNA levels in SKH-1 transfected with either specific RUNX1-EVI-1 siRNA or control siRNA. The graph shows the mean and SEM of 4 independent experiments.

C-D) RUNX1-EVI-1 siRNA results in a decrease of CD34 on the SKH-1 cell surface. Flow cytometric analysis of SKH-1 cells co-staining with CD117-APC and CD34-PE: untreated, or after 14 days with either electroporation alone or control siRNA or RUNX1-EVI-1 siRNA transfection. Gating based on CD34+CD117+ untreated SKH-1 cells. C) Representative flow cytometry plots (CD34-PE versus CD117-APC) D) percentage of CD34+CD117+ cells. Mean of 6 independent experiments and error bars represent S.E.M. * denotes $p<0.05$ by paired t-test.

E-F) RUNX1-EVI-1 siRNA results in down regulation of *MEIS1*, *GATA2* and *HOXA9* but upregulation of *MPO*, *CSF1R* and *CTSG*; as compared to control siRNA treatment alone. mRNA levels in SKH-1 cells following treatment with RUNX1-EVI-1 siRNA as compared to control siRNA transfection, relative to *GAPDH* expression. mRNA levels relative to those from untreated SKH-1. E) mRNA levels of genes required for haematopoietic stem cell function (*GATA2*, *MEIS1* and *HOXA9*) and F) genes expressed after myeloid differentiation (*MPO*, *CTSG* and *CSF1R*). The graph displays mean and SEM of 4 independent experiments.

G) CD14 levels on SKH-1 cells do not increase after RUNX1-EVI-1 siRNA. Flow cytometric analysis of t(3;21) SKH-1 cells stained with CD14-FITC: SKH-1 either untreated, or with 14 days of either electroporation alone, or control siRNA or RUNX1-EVI-1 siRNA transfection. MFI (median) of CD14 FITC on treated cells relative to untreated SKH-1. Graph of mean and SEM of 6 independent experiments.

3.4.1 Effects of RUNX1-EVI-1 siRNA are specific to SKH-1 and not K562 cells

K562 and SKH-1 cells are both derived from patients with CML in blast crisis (Mitani et al., 1994, Lozzio and Lozzio, 1975) and express the BCR-ABL fusion protein. To ensure that the effects of RUNX1-EVI-1 siRNA on the differentiation and apoptosis of the SKH-1 cells are specific to RUNX1-EVI-1 expressing cells we repeated these siRNA transfection experiments in the K562 cell line. Transfection with RUNX1-EVI-1 siRNA resulted in an increase in apoptosis in

the t(3;21) SKH-1 cells but not K562 cell line (figure 3-10A-B). RUNX1-EVI-1 siRNA also specifically reduced the growth of SKH-1 cells but did not affect the growth of K562 cells (figure 3-10C). Finally, in contrast to figure 3-9E-F, in K562 cells there were no changes in expression of key genes involved in haematopoiesis, following 14 days of treatment with RUNX1-EVI-1 siRNA as compared to control siRNA transfection. Therefore, we can conclude that the differentiation and apoptosis of SKH-1 cells, following RUNX1-EVI-1 siRNA transfection, is a result of the knockdown of RUNX1-EVI-1, and not to non-specific effects related to the experimental process.

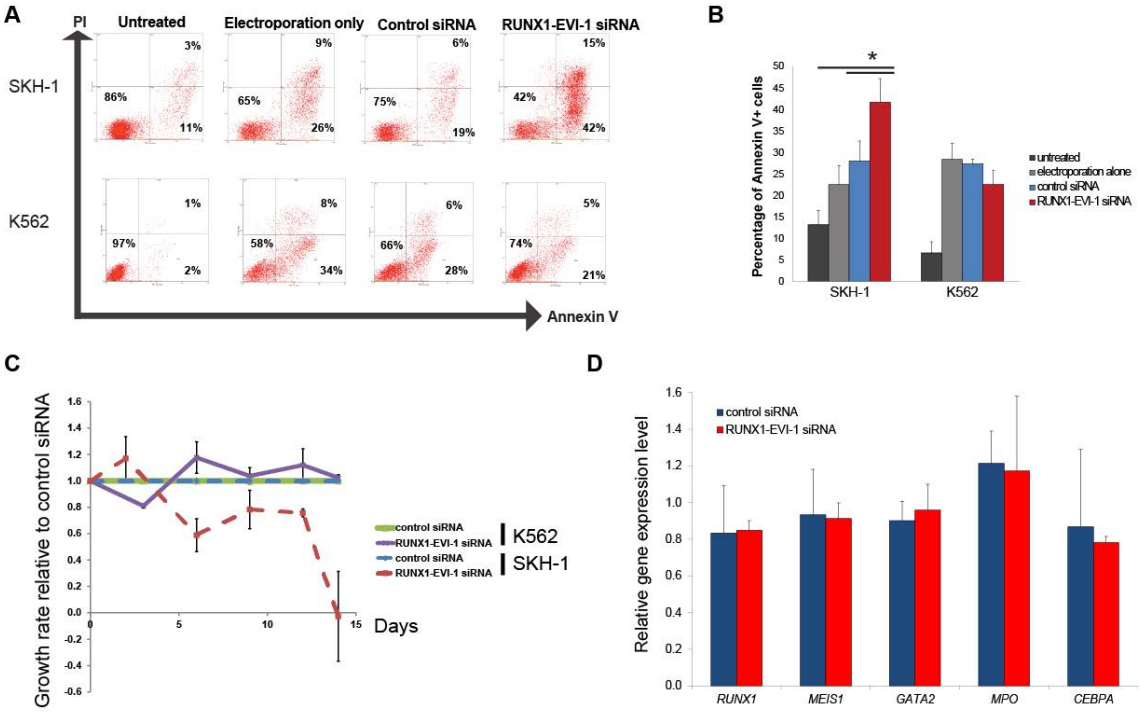


Figure 3-10 The effects of RUNX1-EVI-1 siRNA is specific to t(3;21) SKH-1 cells

A –C) Knockdown of RUNX1-EVI-1 results in reduced cell growth and increased apoptosis of SKH-1 but not K562 cells. Annexin V FITC and PI staining in SKH-1 and K562 cells either untreated or after 14 days with electroporation alone or, either, control siRNA or RUNX1-EVI-1 siRNA transfection. A) Representative plots showing Annexin V-FITC and PI signals. B) Percentage of Annexin V stained cells. The graph shows mean and SEM of at least 3 independent experiments. * denotes $p<0.05$ by unpaired t-test. C) Growth rates of SKH-1 (dashed lines) and K562 cells (solid lines) treated with RUNX1-EVI-1 siRNA relative to treatment with control siRNA. The graph shows mean and SEM of at least 3 independent experiments.

D) Effects of RUNX1-EVI-1 knock-down on expression of genes involved in haematopoiesis are specific to SKH-1 and not K562. Graph shows mRNA levels of genes expressed after myeloid differentiation (*MPO* and *CEBPA*) and required for haematopoietic stem cell function (*GATA2* and *MEIS1*), relative to *GAPDH* and normalised to untreated cells in K562 cells after RUNX1-EVI-1 siRNA as compared to control siRNA transfection. The graph shows the mean and SEM of 3 independent experiments.

3.5 RUNX1-EVI-1 directly regulates key regulators of leukaemic cell identity

Having shown that RUNX1-EVI-1 is required for the maintenance of the leukaemic phenotype, we wished to identify which genes are directly regulated by RUNX1-EVI-1.

In order to identify gene expression changes following RUNX1-EVI-1 knockdown we analysed RNA-seq of SKH-1 cells following either RUNX1-EVI-1 siRNA transfection, or control siRNA transfection. We then used RUNX1-EVI-1 ChIP-seq to identify targets of the fusion protein that change expression.

3.5.1 RUNX1-EVI-1 knockdown results in loss of a stem cell gene expression program

Genome wide changes in gene expression following 2, 4 and 10 days after RUNX1-EVI-1 or control siRNA transfection of SKH-1 cells were characterised by RNA-seq. The aligned reads of RNA-seq are presented in table 3-4. Figure 3-11A shows hierarchical clustering of RNA-seq signals for two independent knock-down time courses demonstrating a strong concordance between the replicates. Clustering also shows that after 4 or 10 days treatment, RNA-seq data in SKH-1 cells treated with RUNX1-EVI-1 siRNA cluster separately to SKH-1 cells treated with control siRNA alone. However, after 2 days of siRNA treatment alone, both control siRNA and RUNX1-EVI-1 siRNA treated RNA-seq samples form a cluster apart from the two latter time points. This suggests that changes in gene expression in SKH-1 cells after 4 or 10 days of RUNX1-EVI-1 siRNA as compared to control siRNA treatment is greater than after 2 days of

RUNX1-EVI-1 siRNA transfection. This is confirmed by figure 3-11B which is a bar graph that show the number of genes that ether increase or decrease in expression (at least 1.5 fold difference between RUNX1-EVI-1 siRNA and control siRNA treatment). Overall, more genes were downregulated than upregulated. The number of genes that responded progressively increased in number over the course of time. By 10 days after RUNX1-EVI-1knockdown, 180 genes were upregulated and 339 genes were downregulated. The heatmap in figure 3-11C depicts all the genes that changed expression in RUNX1-EVI-1 siRNA transfected cells as compared to control siRNA transfection alone, and clusters the RNA-seq data from each time point after 2, 4 or 10 days of siRNA treatment in an unsupervised fashion. Figure 3-11C shows that genes changing expression do so with increasing magnitude as the time course progresses, which is depicted by the increasing intensity of the colours in this heatmap. This timespan suggests that changes in regulation of certain genes may require time, in order to take place. Alternatively, the changes of gene expression at later time points may also be a result of wider changes in the transcriptional network of these cells initiated by changes in RUNX1-EVI-1 binding. Finally, the progressive changes in gene expression over time following RUNX1-EVI-1 knockdown may also be reflective of the gradual degradation of the fusion protein following initial RUNX1-EVI-1 siRNA transfection leading to a cell autonomous response and cellular heterogeneity.

Having manually validated that a number of genes critical for stem cell function were downregulated following RUNX1-EVI-1 knockdown, we wished to identify whether or not these changes were consistent more broadly by a loss of a coordinated stem cell gene expression program. Eppert et al profiled gene expression in, FACS sorted, functionally validated HSCs and LSCs from primary AML samples (Eppert et al., 2011). This gene signature was validated outside of the xenograft model as this HSC and LSC gene set was shown to have prognostic value in large, independent cohorts of AML patients. Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) is a bioinformatics technique that ranks genes by expression levels in one sample and determines statistically, how well represented is an independently defined set of genes.

GSEA analysis demonstrated that both the HSC and LSC gene signature (Eppert et al., 2011) were enriched in the genes that were down-regulated following RUNX1-EVI-1 knockdown (figure 3-11D).

Next, we used KEGG pathway analysis to categorise sets of genes that change expression following RUNX1-EVI-1 knockdown. This analysis maps the large number of genes that are differentially expressed in RUNX1-EVI-1 siRNA transfected cells as compared to control siRNA, to their known roles. This enables a description of common and highly represented functions that change following RUNX1-EVI-1 knockdown. Up-regulated pathways included "Lysosome" (e.g. *CTSG*, *ASAH1*, and *GLA*) and "FcγR mediated phagocytosis" (e.g. *FCGR2A*, *NCF1*, *GAB)* (figure 3-11E) which are functions characteristic of fully differentiated neutrophils and monocytes/macrophages. Down regulated gene classes include "Transcriptional misregulation in cancer" (including *MYCN*, *WT1*, *ID2*, *JUP*, *ERG*, *COMMD3-BMI-1*) and "Cell adhesion molecules" (e.g. *CD34*, *ICAM3, ITGA9*).

Together, this suggests that RUNX1-EVI-1 knockdown results in a loss of a stem cell specific program characterised by changes in expression of epigenetic regulators vital for their function, with an up-regulation of genes required for terminal myeloid function. This is in concordance with the phenotypic changes in SKH-1 cells following RUNX1-EVI-1 knockdown described in chapter 3.4.

Table 3-4 RUNX1-EVI-1 or control siRNA transfected SKH-1 RNA-seq reads alignment
Control siRNA (siMM) or RUNX1-EVI-1 siRNA (siREVI1) transfected cells. RNA –seq in SKH-1 cells after either 2 (D2), 4 (D4) or 10 (D10) days of siRNA treatment. Two biological replicates performed (Rep1 and Rep2).

RNA-seq Dataset	Aligned reads (n)
siMM_D2_Rep1	43584690
siMM_D2_Rep2	41344409
siMM_D4_Rep1	33410586
siMM_D4_Rep2	44832782
siMM_D10_Rep1	22705489
siMM_D10_Rep2	22419529
siREVI1_D2_Rep1	43782238

siREVI1_D2_Rep2	23157018
siREVI1_D4_Rep1	37541312
siREVI1_D4_Rep2	53187360
siREVI1_D10_Rep1	17656510
siREVI1_D10_Rep2	30380957

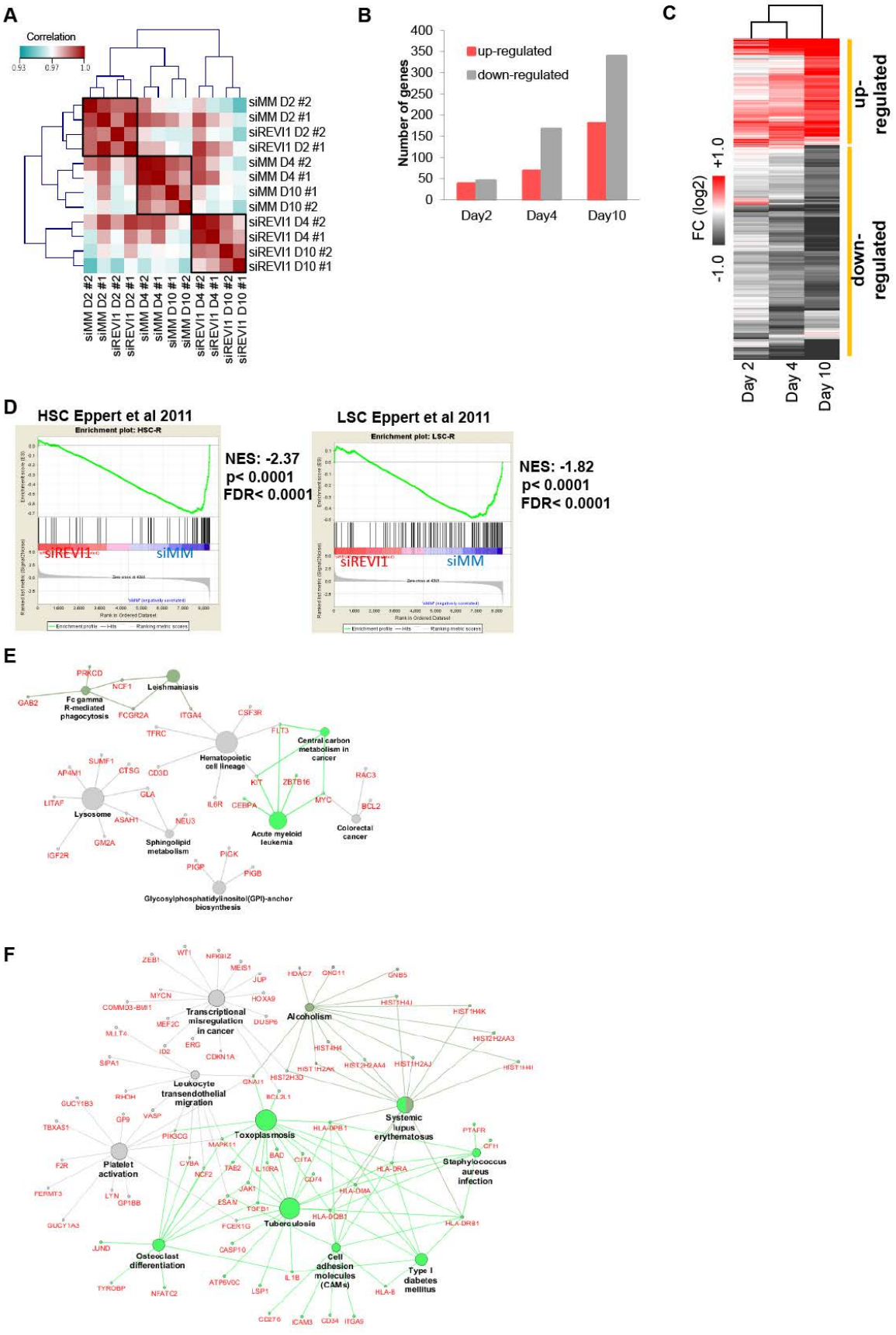

A
Correlation
0.93
0.97
1.0
siMM D2 #2
siMM D2 #1
siREVI1 D2 #2
siREVI1 D2 #1
siMM D4 #2
siMM D4 #1
siMM D10 #1
siMM D10 #2
siREVI1 D4 #2
siREVI1 D4 #1
siREVI1 D10 #2
siREVI1 D10 #1
B
Number of genes
up-regulated
down-regulated
Day2
Day4
Day10
C
FC (log2)
+1.0
-1.0
up-regulated
down-regulated
Day 2
Day 4
Day 10
D
HSC Eppert et al 2011
Enrichment plot: HSC-R
siREVI1
siMM
NES: -2.37
p< 0.0001
FDR< 0.0001
LSC Eppert et al 2011
Enrichment plot: LSC-R
siREVI1
siMM
NES: -1.82
p< 0.0001
FDR< 0.0001
E
PRKCD
NCF1
Leishmaniasis
GAB2
Fc gamma R-mediated phagocytosis
FCGR2A
ITGA4
CSF3R
TFRC
FLT3
Central carbon metabolism in cancer
Hematopoietic cell lineage
SUMF1
AP4M1
CTSG
CD3D
RAC3
KIT
ZBTB16
LITAF
GLA
Lysosome
IL6R
CEBPA
MYC
BCL2
ASAH1
NEU3
Acute myeloid leukemia
Colorectal cancer
IGF2R
GM2A
Sphingolipid metabolism
PIGK
PIGP
PIGB
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis
F
ZEB1
WT1
NFKBIZ
MEIS1
JUP
HDAC7
GNG11
GNB5
MYCN
HOXA9
COMMD3-BMI1
Transcriptional misregulation in cancer
DUSP6
Alcoholism
HIST1H4J
HIST1H4K
MEF2C
MLLT4
ERG
HIST2H2AA3
ID2
CDKN1A
HIST4H4
SIPA1
HIST1H2AK
HIST2H2AA4
HIST1H2AJ
HIST1H4I
GNAI1
HIST2H3D
Leukocyte transendothelial migration
RHOH
BCL2L1
GUCY1B3
GP9
VASP
Systemic lupus erythematosus
TBXAS1
PIK3CG
Toxoplasmosis
HLA-DPB1
PTAFR
MAPK11
CFH
F2R
CYBA
BAD
CITA
Platelet activation
NCF2
TAE2
IL10RA
HLA-DRA
Staphylococcus aureus infection
CD74
FERMT3
JAK1
HLA-DMA
LYN
GP1BB
ESAM
TGFB1
Tuberculosis
HLA-DQB1
GUCY1A3
FCER1G
HLA-DR81
JUND
CASP10
Osteoclast differentiation
Cell adhesion molecules (CAMs)
Type I diabetes mellitus
TYROBP
ATP6V0C
IL1B
NFATC2
LSP1
HLA-E
CD276
ICAM3
CD34
ITGA5

Figure 3-11 Progressive changes in gene expression following RUNX1-EVI-1 knockdown results in a loss of stem cell gene expression program
RNA-seq analysis of SKH-1 cells after either RUNX1-EVI-1 or control siRNA treatment for 2, 4 or 10 days. Biological replicates at each time point and with each treatment condition are indicated.

A) Hierarchical clustering of Pearson correlation coefficients between two biological replicates (#1 and #2) of each treatment condition (siMM: control siRNA; siREVI1: RUNX1-EVI-1 siRNA). Independent replicates cluster together but also three major clusters: samples from day 2 timepoint (both control siRNA and RUNX1-EVI-1 siRNA treated samples), samples from control siRNA treated samples at day 4 and 10, and samples from RUNX1-EVI-1 siRNA treated samples at day 4 and day 10.

B) Bar plot shows number of down-regulated genes and up-regulated genes at each timepoint (2, 4, or 10 days of treatment). Numbers of both up-regulated and down-regulated genes increase as the duration of siRNA transfection lengthens. Values based on the average RNA-seq signal in two independent replicates in SKH-1 cells transfected with either RUNX1-EVI-1 siRNA or control siRNA. This is used to determine the number of differentially expressed genes in SKH-1 cells that change expression 1.5 fold after RUNX1-EVI-1 siRNA as compared to control siRNA transfection.

C) Samples at different time points of treatment clustered by RNA-seq. SKH-1 cells treated for 10 days with RUNX1-EVI-1 siRNA cluster away from those treated for 2 or 4 days alone. Heat map of genes which change expression 1.5 fold after RUNX1-EVI-1 siRNA transfection, as compared to control siRNA. The graph shows the average of two independent replicates. Heat map colour intensity is related to the degree of differential expression (fold-change (FC)) between RUNX1-EVI-1 siRNA and control siRNA treatment.

D) Gene set enrichment analysis (GSEA) of RNA-seq from SKH-1 cells after 10 days of treatment with either RUNX1-EVI-1 siRNA (siREVI1) or control siRNA (siMM). Loss of enrichment of genes associated with either haematopoietic or leukaemic stem cell when SKH-1 are treated with RUNX1-EVI-1 siRNA.

E) KEGG pathway for genes up-regulated in expression in SKH-1 cells treated after 10 days of RUNX1-EVI-1 siRNA as compared to control siRNA.

F) KEGG pathway for genes downregulated in expression in SKH-1 cells treated after 10 days of RUNX1-EVI-1 siRNA as compared to control siRNA.

3.5.2 RUNX1-EVI-1 directly targets key genes involved in differentiation and cell survival.

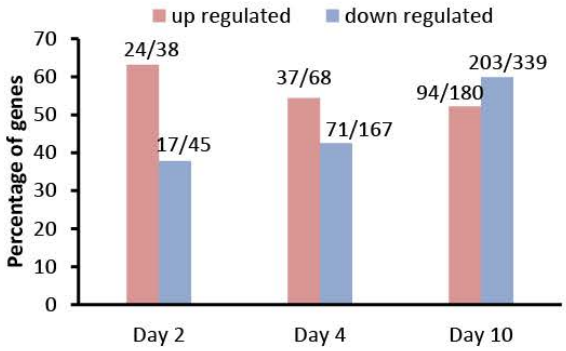
By integrating ChIP-seq of RUNX1-EVI-1 with changes in gene expression following RUNX1-EVI-1 knockdown we can identify which genes are directly

regulated by RUNX1-EVI-1 and whether RUNX1-EVI-1 acts as an activator and repressor of genes.

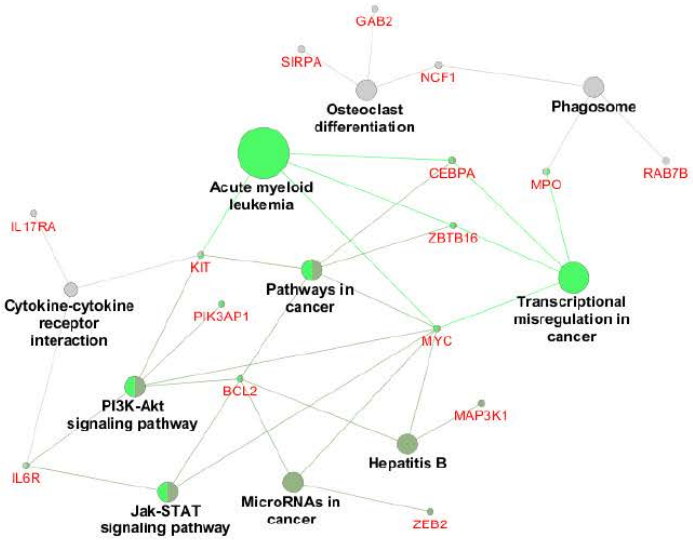
Figure 3-12A show that between 40-50% of genes which change expression by day 10 are direct targets of RUNX1-EVI-1. Targets of RUNX1-EVI-1 included both up and down regulated genes following RUNX1-EVI-1 knockdown. Changes in gene expression earlier in the time course are more likely to be a direct result of changes to RUNX1-EVI-1 binding, as later events may be influenced by indirect changes, for example, downstream of a RUNX1-EVI-1 target which in itself is a regulator. At day 2 after RUNX1-EVI-1 knockdown, more genes were up-regulated, than down regulated and a greater proportion of the total set of genes which are up-regulated were RUNX1-EVI-1 targets. Overall, this suggests that although RUNX1-EVI-1 may act as a transcriptional repressor, it may also lead to the activation of other loci.

Having noted the large proportion of genes that changed expression are RUNX1-EVI-1 targets, it was not surprising to note that KEGG pathway analysis of RUNX1-EVI-1 targets that change expression bear a strong similarity to those already shown in figure 3-11. Figure 3-12B demonstrates that RUNX1-EVI-1 appears to directly repress genes that are required in granulocytes and monocytes, such as genes important for phagocytosis and the key transcription factor *CEBPA.* Simultaneously RUNX1-EVI-1 may directly or indirectly activate genes involved in “transcriptional mis-regulation in cancer” such as *ERG* and *WT1*, as well as *TGFB1.* These genes are involved in the maintenance of stem cell properties, and are often deregulated in cancerous cells. For example, *WT1* is involved in the regulation of quiescence in haematopoietic stem cells (Ellisen et al., 2001) and is over-expressed in poor prognosis AML (Rampal and Figueroa, 2016).

A


up regulated
down regulated
Percentage of genes
70
60
50
40
30
20
10
0
24/38
17/45
37/68
71/167
94/180
203/339
Day 2
Day 4
Day 10

B


GAB2
SIRPA
NCF1
Osteoclast differentiation
Phagosome
Acute myeloid leukemia
CEBPA
MPO
RAB7B
IL17RA
ZBTB16
KIT
Pathways in cancer
Cytokine-cytokine receptor interaction
PIK3AP1
Transcriptional misregulation in cancer
MYC
PI3K-Akt signaling pathway
BCL2
MAP3K1
Hepatitis B
IL6R
Jak-STAT signaling pathway
MicroRNAs in cancer
ZEB2

C

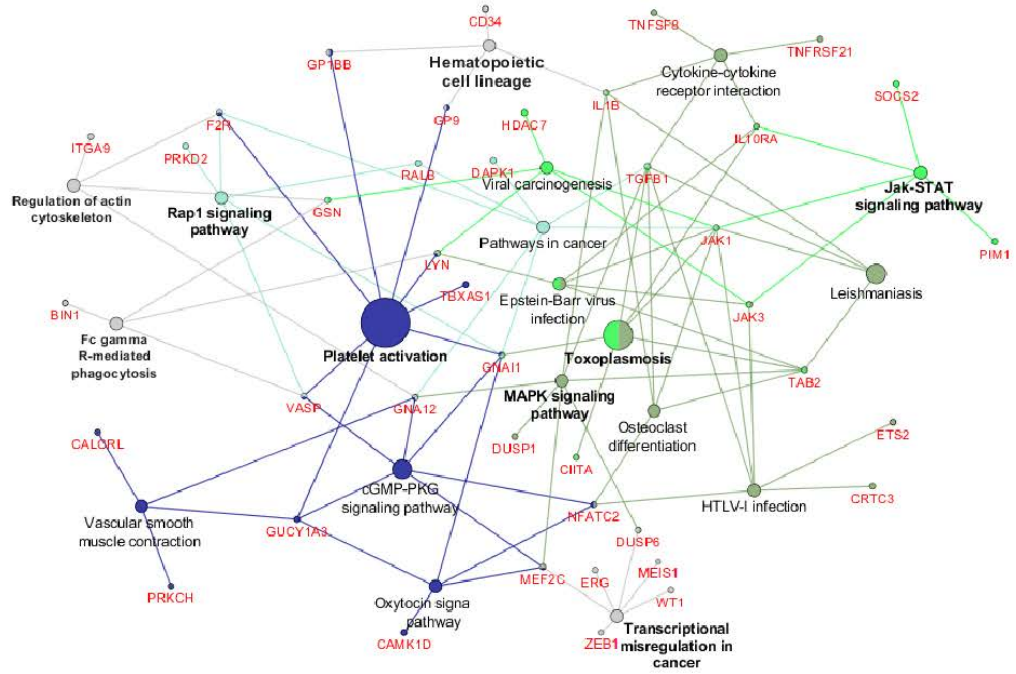

CD34
TNFSF8
TNFRSF21
GP1BB
Hematopoietic cell lineage
Cytokine-cytokine receptor interaction
SOCS2
IL1B
F2R
GP9
HDAC7
ITGA9
IL10RA
PRKD2
RALB
DAPK1
Viral carcinogenesis
TGFB1
Jak-STAT signaling pathway
Regulation of actin cytoskeleton
Rap1 signaling pathway
GSN
Pathways in cancer
JAK1
PIM1
LYN
Leishmaniasis
TBXAS1
Epstein-Barr virus infection
BIN1
JAK3
Fc gamma R-mediated phagocytosis
Platelet activation
Toxoplasmosis
GNAI1
TAB2
MAPK signaling pathway
VASP
GNA12
Osteoclast differentiation
ETS2
CALCRL
DUSP1
CIITA
cGMP-PKG signaling pathway
HTLV-I infection
CRTC3
Vascular smooth muscle contraction
NFATC2
GUCY1A3
DUSP6
MEIS1
MEF2C
ERG
PRKCH
WT1
Oxytocin signa pathway
Transcriptional misregulation in cancer
CAMK1D
ZEB1

Figure 3-12 Identification of RUNX1-EVI-1 targets responding to siRNA treatment
A) Percentage of differentially expressed genes which are RUNX1-EVI-1 ChIP-seq targets. The fraction displayed on top of columns represents absolute number of differentially expressed genes shown as: RUNX1-EVI-1 targets/total number of differentially expressed genes. Differentially expressed genes are those which have at 1.5 fold change in gene expression between RUNX1-EVI-1 siRNA as compared to control siRNA treatment, either after 2, 4 or 10 days of siRNA treatment.

B)-C) RUNX1-EVI-1 expression leads to both activation of leukaemia pathways as well as the repression of terminal myeloid differentiation. KEGG pathway for RUNX1-EVI-1 ChIP-seq target genes which are at least 1.5 fold differentially expressed, between RUNX1-EVI-1 siRNA and control siRNA treated cells (after 10 days of treatment). B) Pathways upregulated following RUNX1-EVI-1 knockdown and C) pathways downregulated following RUNX1-EVI-1 knockdown.

3.6 C/EBPα remodels the epigenome of SKH-1 cells following RUNX1-EVI-1 knockdown

3.6.1 DHS mapping in SKH-1 following RUNX1-EVI-1 knockdown increases DHS containing CEBP motifs.

Having identified the changes in gene expression in SKH-1 cells following RUNX1-EVI-1 knockdown, we characterised the alterations in the epigenome driving these changes. DHSs are enriched for gene regulatory elements such as enhancers (Cockerill, 2011). In order to identify changes in these gene regulatory elements we performed DNase-seq in SKH-1 cells following either 2 or 10 days of either control or RUNX1-EVI-1 siRNA transfection. The sequencing parameters of these experiments are presented in table 3-5 below.

Table 3-5 DNase seq in SKH-1 after 2 or 10 days treatment of control or RUNX1-EVI-1 siRNA

DNase-seq Dataset	Aligned reads (n)	Peaks (n)
Control siRNA Day 2	17,680,722	21,261
RUNX1-EVI-1 siRNA Day 2	22,242,583	22,312
Control siRNA Day 10	8,531,753	16,456

RUNX1-EVI-1 siRNA Day 10	19,976,087	18,672

After 2 days of RUNX1-EVI-1 siRNA treatment, there were relatively few genes change expression and hence relatively few DHSs change in their accessibility (figure 3-11C and 3-13A). After 10 days of RUNX1-EVI-1 siRNA treatment, a greater number of genes changed expression and thus a greater number of DHSs changed in their accessibility. Figure 3-13B and C depict changes in DHSs following RUNX1-EVI-1 knockdown. For example, after 10 days of siRNA treatment (Figure 3-13D) 294 DHSs were reduced in hypersensitivity, the majority remained unchanged but 2510 DHSs were increased in hypersensitivity, following RUNX1-EVI-1 knockdown. However, already after 2 days as well as after 10 days of knock-down, sites increasing in hypersensitivity following RUNX1-EVI-1 were enriched for CEBP motifs in a de novo motif analysis. This suggests an increased role for the CEBP family of transcription factors in SKH-1 cells following knockdown of RUNX1-EVI-1. This finding is reflective of our findings for the importance of C/EBPα in driving differentiation of t(8;21) cells, following RUNX1-ETO knockdown (Ptasinska et al., 2014).

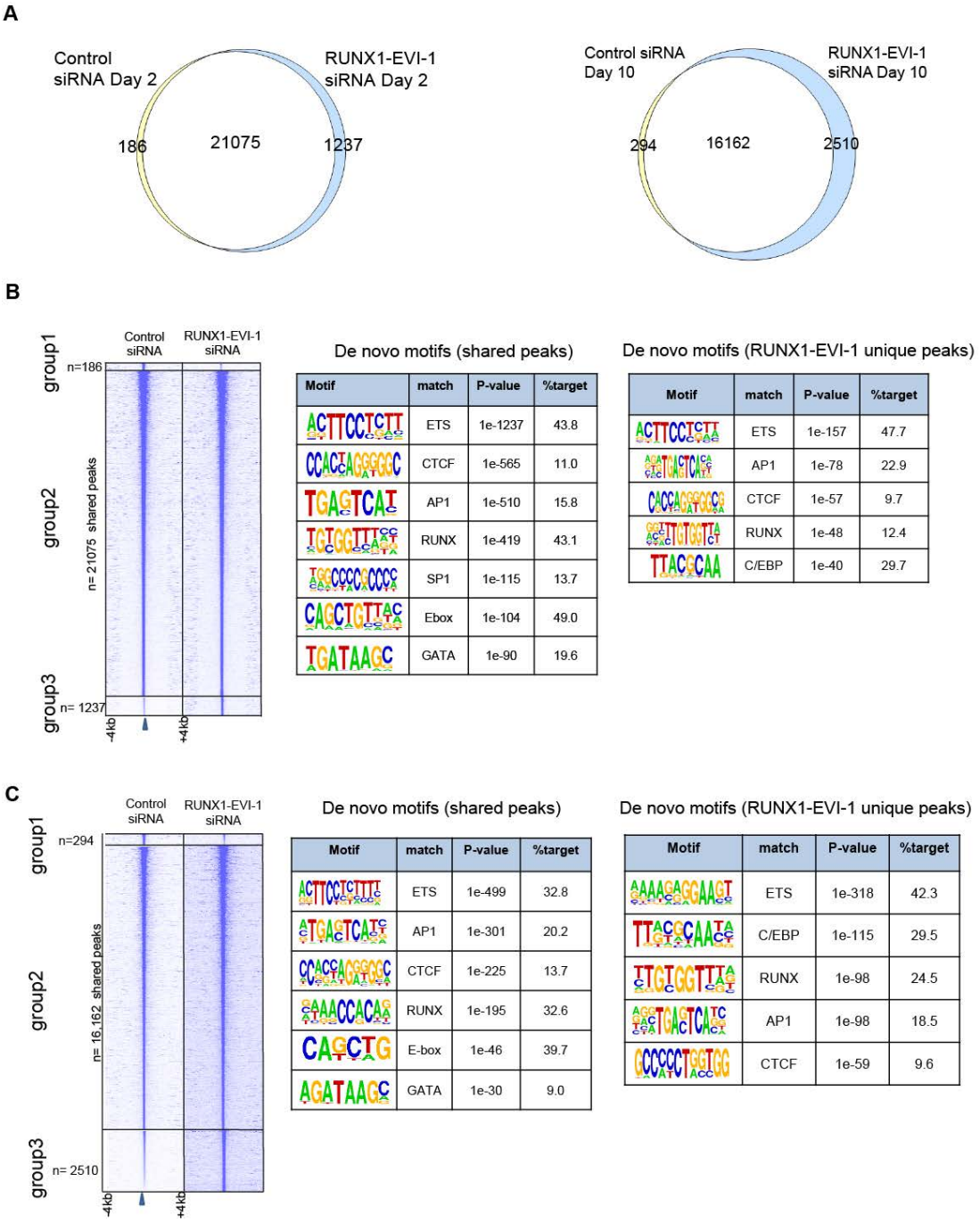


Figure 3-13 RUNX1-EVI-1 knockdown results in increase in DHSs with C/EBP motifs

DNase-seq in SKH-1 after 2 or 10 days of either control or RUNX1-EVI-1 siRNA treatment.

A) Identification of DHSs found only in SKH-1 treated with control siRNA, only in SKH-1 treated with RUNX1-EVI-1 siRNA, or found in SKH-1 treated with either

siRNA. Venn diagram of peak overlaps after either 2 or 10 days of either control or RUNX1-EVI-1 siRNA treatment.

B-C) Enrichment of CEBP motif containing DHSs in SKH-1 following transfection of RUNX1-EVI-1 specific siRNA. DNase-seq tag density within +/- 4Kb window centred on DNase-seq peaks in SKH-1 treated with B) 2 days or C) 10 days of control siRNA. To the right, aligned to the same coordinates are DNase-seq tag densities for SKH-1 treated with RUNX1-EVI-1 siRNA.

Tables showing de novo motif discovery in DHS present in both control and RUNX1-EVI-1 siRNA treated cells (labelled shared peaks) or only in RUNX1-EVI-1 siRNA treated cells (labelled RUNX1-EVI-1 unique peaks) are shown on the right.

To determine the significance of the changes in chromatin accessibility after knock-down, we sought to understand whether they were associated with changes in gene expression of their neighbouring genes by subdividing them into three groups. The majority of gene expression changes occurred at DHSs that did not change (group 2, figure 3-14A). However, group 1 DHSs that had decreased in hypersensitivity following RUNX1-EVI-1 knockdown show a decrease in expression of neighbouring genes. Group 3 DHSs that increase in hypersensitivity following RUNX1-EVI-1 siRNA, showed an increase in expression of neighbouring genes at the initial time point.

To assess which motifs were specifically enriched in any of these three groups of DHSs, we plotted the underlying motif frequencies at the same DHS loci (figure 3-14B). This demonstrates an enrichment of GATA motif containing DHSs in group 1. This is consistent with a decrease in expression of *GATA2* following RUNX1-EVI-1 knockdown in SKH-1 cells (figure 3-9E)

In contrast, DHSs increasing in accessibility after RUNX1-EVI-1 knockdown showed and enrichment of CEBP motifs (group 3). RUNX motifs containing DHSs also increase in accessibility following RUNX1-EVI-1 knockdown (figure 3-14B) suggesting a potential cooperativity between RUNX and CEBP transcription factors. Although RUNX1 expression was unchanged (figure 3-9A), cooperativity between RUNX1 and CEBP transcription factors in binding

DNA has previously been documented (Zhang et al., 1996, Petrovick et al., 1998). In this study, we characterise this further in figure 3-17, 3-18 and 3-21

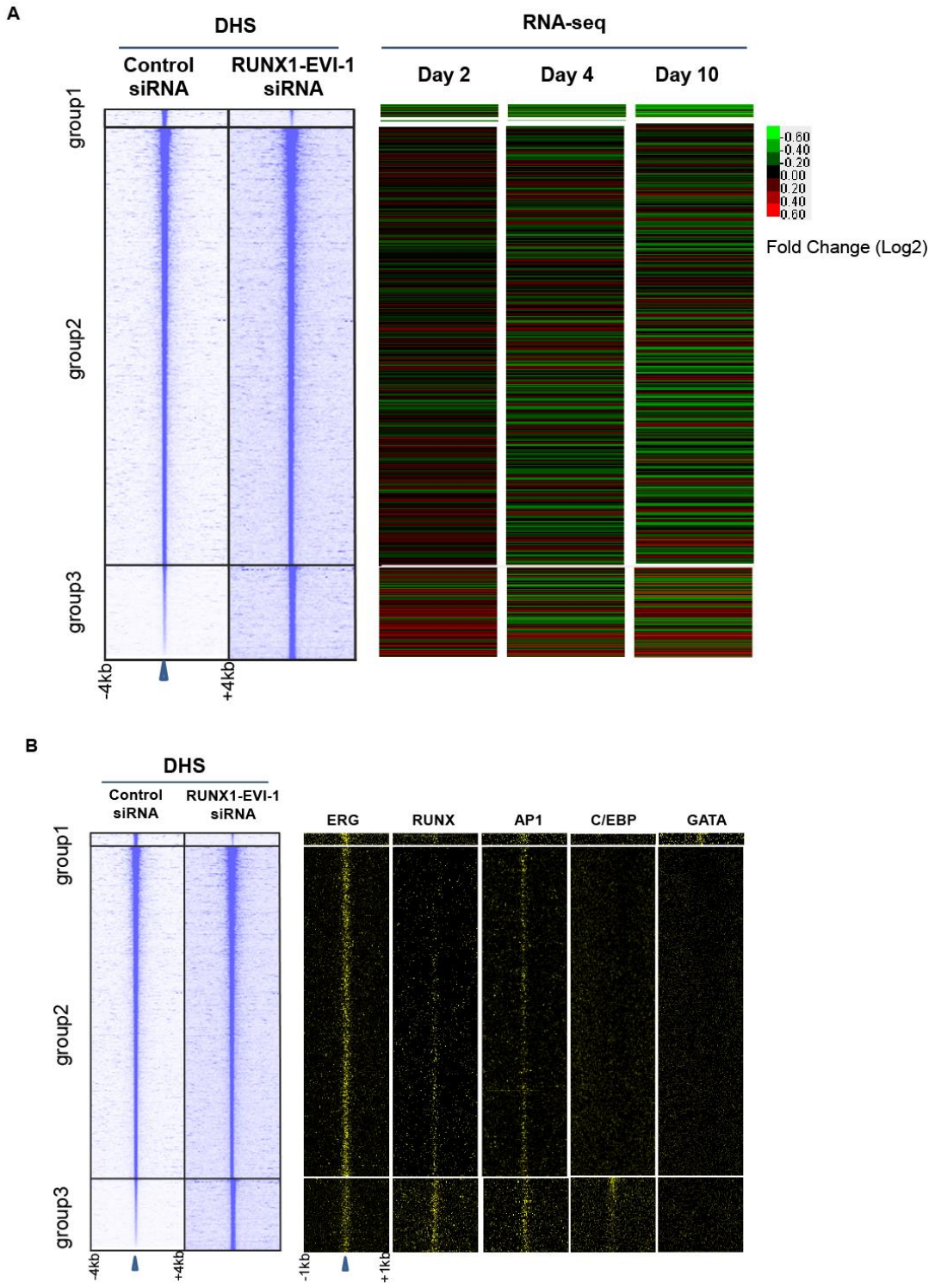


Figure 3-14 Changes in DHS accessibility by RUNX1-EVI-1 siRNA correlate with changes in the expression of nearby genes

Profiles of the DNase-seq tag densities, in +/-4Kb windows centred on peaks, for SKH-1 treated for 10 days with either control or RUNX1-EVI-1 siRNA. Group

1 peaks are DHSs which decrease in accessibility following RUNX1-EVI-1 knockdown. Group 2 peaks are DHSs whose accessibility is unchanged following RUNX1-EVI-1 knockdown. Group 3 peaks are DHS which increase in accessibility following RUNX1-EVI-1 knockdown.

A) Group 1 DHSs are largely associated with decrease in expression of neighbouring genes. Group 3 DHSs are associated with increase in expression of neighbouring genes after 2 days of siRNA treatment. Gene expression levels of associated genes at the same coordinates are shown to the right, depicted as RUNX1-EVI-1 siRNA log2 fold change, after 2, 4 or 10 days of siRNA treatment as compared to control RNA treatment.

B) Motif content of Group 1 – 3 DHSs showing that GATA motif containing DHSs decrease in accessibility following RUNX1-EVI-1 knockdown whereas RUNX1 and CEBP motif containing DHSs increase in accessibility following RUNX1-EVI-1 knockdown. Motif densities are plotted to the right within +/-1 Kb windows) of the DHSs centre.

3.6.2 CBF fusion genes in both t(3;21) and t(8;21) leukaemia deregulates C/EBPα expression

We next investigated which CEBP transcription factor might bind DHSs which increasing in accessibility in SKH-1 following RUNX1-EVI-1 knockdown (figure 3-14A). *CEBPA* was prominent as it was a direct target of RUNX1-EVI-1 (figure 3-12). Manual RT-qPCR and Western blotting confirmed that C/EBPα mRNA and protein levels increased after RUNX1-EVI-1 knockdown (Figure 3-15A-B). We checked the RNA levels of all six members of the CEBP transcription factor family (Figure 3-15C) following RUNX1-EVI-1 knockdown. This showed that *CEBPA* and *CEBPD* RNA both increased in SKH-1 cells transfected with RUNX1-EVI-1 siRNA as compared to control siRNA. However, *CEBPA* RNA levels were approximately 3 fold higher than *CEBPD*. *CEBPE* RNA levels remained low throughout the time course of RUNX1-EVI-1 knockdown. This is of interest because in t(8;21) cells following RUNX1-ETO knockdown, *CEBPA*, *CEBPD* and *CEBPE* RNA levels also increase (Ptasinska et al., 2012, Ptasinska et al., 2014). *CEBPE* expression is principally required for the terminal development and appearance of neutrophils (Yamanaka et al., 1997), although some defects in macrophage function is seen in *Cebpe* null mice (Tavor et al., 2002). This suggests that the differentiation of t(3;21) cells

remains incomplete, following knockdown of RUNX1-EVI-1, as compared to t(8;21) cells, following knockdown of RUNX1-ETO.

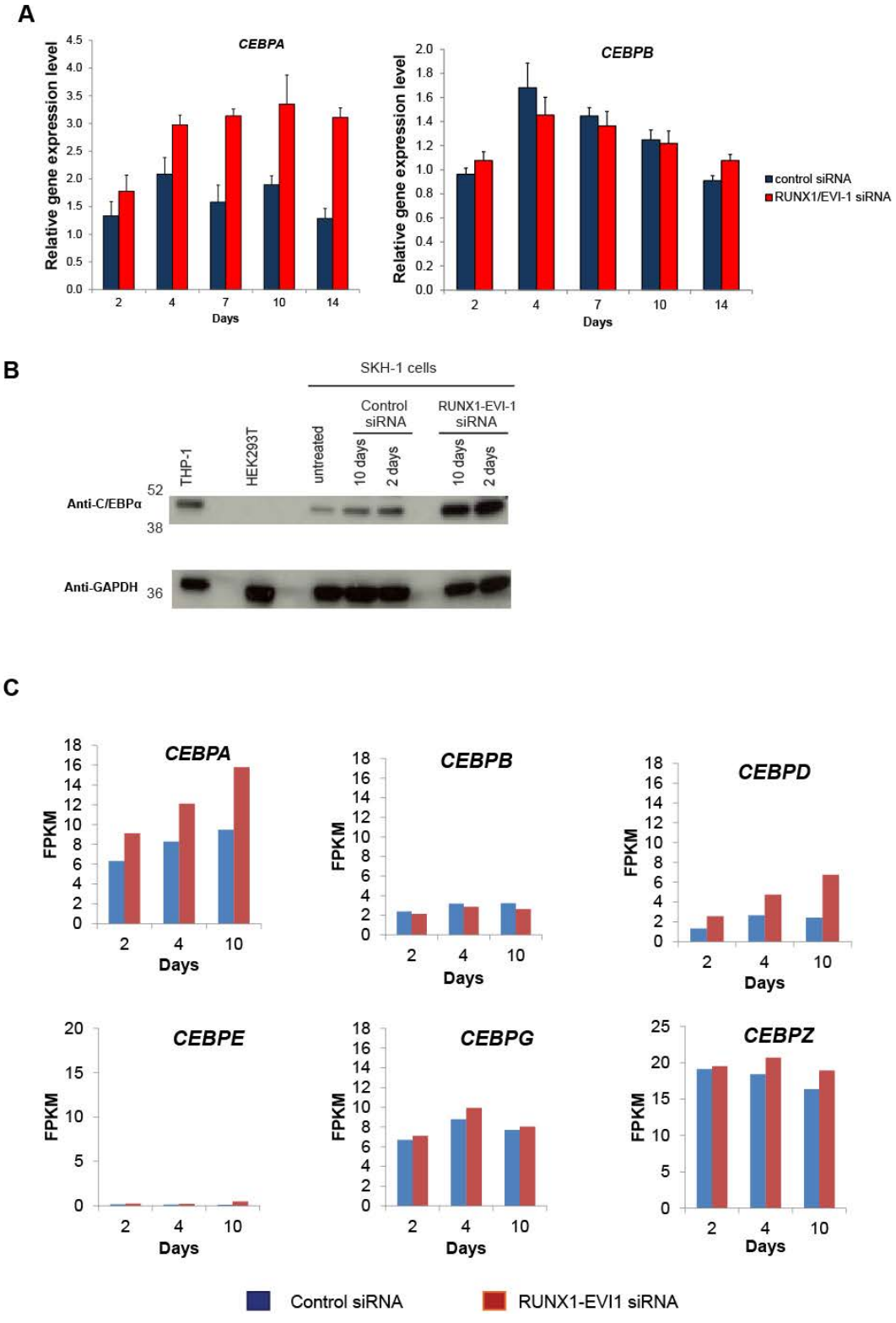


Figure 3-15 CEBPA expression increases after RUNX1-EVI-1 knockdown
A) *CEBPA* but not *CEBPB* mRNA levels increase following knockdown of RUNX1-EVI-1 in t(3;21) cells. RT-qPCR of mRNA levels relative to *GAPDH*

and normalised to untreated cells of *CEBPA* and *CEBPB* in SKH-1 after RUNX1-EVI-1 siRNA as compared to control siRNA transfection. Cells harvested 2, 4, 7, 10 or 14 days after start of time course. The graph shows the mean and SEM of 4 independent experiments.

B) C/EBPα protein levels increase in SKH-1 cells following knockdown of RUNX1-EVI-1. Western blot from whole cell lysates in t(3;21) SKH-1 cell treated with control siRNA or RUNX1-EVI-1 siRNA after 2 and 10 days probed with an anti-C/EBPα antibody. Whole cell lysates from HEK293T and THP-1 were used as negative control and positive control, respectively. The GAPDH signal served as loading control.

C) Expression of C/EBP transcription factors after RUNX1-EVI-1 knockdown. Increase in *CEBPA* and *CEBPD* RNA levels following RUNX1-EVI-1 knockdown in SKH-1 cells. *CEBPA*, *CEBPB*, *CEBPD*, *CEBPE*, *CEBPZ* and *CEBPG* FPKM from RNA-seq in SKH-1 cells after 2, 4 and 10 days treatment with either control siRNA or RUNX1-EVI-1 siRNA.

CEBPA is also upregulated following RUNX1—ETO knockdown in t(8;21) Kasumi-1 cells (Ptasinska et al., 2014). We therefore asked whether *CEBPA* was commonly down regulated in CBF fusion AML in general using publically available RNA-seq data from an independent cohort of patients (Cancer Genome Atlas Research, 2013). A comparison of normalised RSEM values from RNA-seq in the normal karyotype AML vs CBF fusion leukaemia (seven t(8;21) and one t(3;21) AML) shows that *CEBPA* is expressed at a significantly lower level in CBF fusion leukaemia than normal karyotype leukaemia (figure3-16A). Similarly in another independent cohort of AML patients (Verhaak et al., 2009), *CEBPA* is significantly down-regulated in CBF fusion leukaemia as compared to normal karyotype AML. Of note, in both cohorts, t(3;21) AML *CEPBA* expression is close to the median of the rest of the t(8;21) AML *CEBPA* expression.

Furthermore, *CEBPA* mutations are mutually exclusive from t(8;21) leukaemias suggesting they act on the same pathways: 13 *CEBPA* mutant AML are independent of 8 CBF fusion leukaemia (Cancer Genome Atlas Research, 2013); 66 *CEBPA* biallelic mutant AML are independent of 60 t(8;21) AML

(Papaemmanuil et al., 2016); 9 t(3;21) AML are without CEBPA mutations (H Dohner, personal correspondence).

To further examine whether gene expression changes following RUNX1-EVI-1 and RUNX1-ETO knockdown involve the same developmental pathways we clustered GO terms in genes responding to knockdown of either RUNX1-EVI-1 or RUNX1-ETO. Figure 3-16B demonstrate that despite the differences between the two types of AML, changing gene expression patterns both converge on the same pathways following CBF fusion protein knockdown. GO terms that are shared include “myeloid leukocyte mediated immunity”, “myeloid cell differentiation and “leukocyte degranulation”. This suggests that a core-set of genes are reprogrammed following knockdown of RUNX1-EVI-1 in t(3;21) cells and RUNX1-ETO in t(8;21) cells.

To examine the molecular mechanism of this conversion, we used our ChIP-Seq data to see whether RUNX1-ETO and RUNX1-EVI1 bound to the *CEBPA locus.* An enhancer located at +42kb from the promoter of *CEBPA* is an important cis-regulatory element of this gene that is required for myeloid lineage priming (Avellino et al., 2016). Both RUNX1-ETO and RUNX1-EVI-1 bind to this element (figure 3-16C) and may therefore share similar mechanisms of *CEBPA* repression. Despite the difference in epigenomic landscape in t(3;21) and t(8;21) leukaemia, this data suggests that *CEBPA* down-regulation is a hallmark of both CBF fusion protein leukaemia, likely driven through modulation of the same distal enhancer of *CEBPA.*

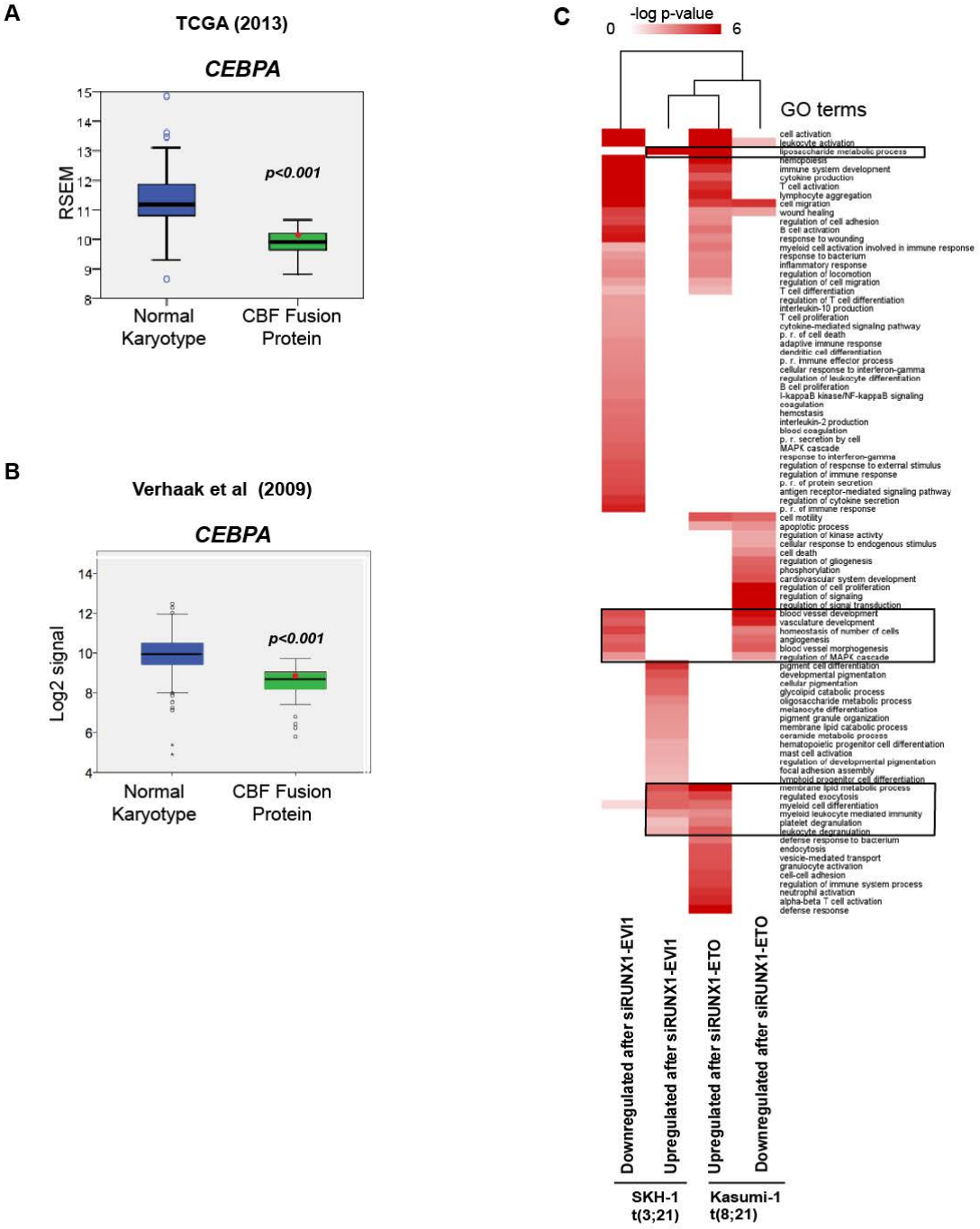

A
TCGA (2013)
CEBPA
RSEM
p<0.001
Normal Karyotype
CBF Fusion Protein
B
Verhaak et al (2009)
CEBPA
Log2 signal
p<0.001
Normal Karyotype
CBF Fusion Protein
C
-log p-value
0
6
GO terms
cell activation
leukocyte activation
lipopsaccharide metabolic process
hemopoiesis
immune system development
cytokine production
T cell activation
lymphocyte aggregation
cell migration
wound healing
regulation of cell adhesion
B cell activation
response to wounding
myeloid cell activation involved in immune response
response to bacterium
inflammatory response
regulation of locomotion
regulation of cell migration
T cell differentiation
regulation of T cell differentiation
interleukin-10 production
T cell proliferation
cytokine-mediated signaling pathway
p. r. of cell death
adaptive immune response
dendritic cell differentiation
p. r. immune effector process
cellular response to interferon-gamma
regulation of leukocyte differentiation
B cell proliferation
I-kappaB kinase/NF-kappaB signaling
coagulation
hemostasis
interleukin-2 production
blood coagulation
p. r. secretion by cell
MAPK cascade
response to interferon-gamma
regulation of response to external stimulus
regulation of immune response
p. r. of protein secretion
antigen receptor-mediated signaling pathway
regulation of cytokine secretion
p. r. of immune response
cell motility
apoptotic process
regulation of kinase activity
cellular response to endogenous stimulus
cell death
regulation of gliogenesis
phosphorylation
cardiovascular system development
regulation of cell proliferation
regulation of signaling
regulation of signal transduction
blood vessel development
vasculature development
homeostasis of number of cells
angiogenesis
blood vessel morphogenesis
regulation of MAPK cascade
pigment cell differentiation
developmental pigmentation
cellular pigmentation
glycolipid catabolic process
oligosaccharide metabolic process
melanocyte differentiation
pigment granule organization
membrane lipid catabolic process
ceramide metabolic process
hematopoietic progenitor cell differentation
mast cell activation
regulation of developmental pigmentation
focal adhesion assembly
lymphoid progenitor cell differentiation
membrane lipid metabolic process
regulated exocytosis
myeloid cell differentiation
myeloid leukocyte mediated immunity
platelet degranulation
leukocyte degranulation
defense response to bacterium
endocytosis
vesicle-mediated transport
granulocyte activation
cell-cell adhesion
regulation of immune system process
neutrophil activation
alpha-beta T cell activation
defense response
Downregulated after siRUNX1-EVI1
Upregulated after siRUNX1-EVI1
Upregulated after siRUNX1-ETO
Downregulated after siRUNX1-ETO
SKH-1 t(3;21)
Kasumi-1 t(8;21)
D
10 kb
hg38
33,265,000
33,270,000
33,275,000
33,280,000
33,285,000
33,290,000
33,295,000
33,300,000
33,305,000
CEBPA
t(3;21)
RUNX1-EVI-1
RUNX1
t(8;21)
RUNX1-ETO
RUNX1
+42Kb enhancer

Figure 3-16 C/EBPα is a common up-regulated target after RUNX1-EVI-1 and RUNX1-ETO knockdown
A-B) *CEBPA* mRNA is down-regulated in CBF leukaemia as compared to normal karyotype AML.

A) RSEM normalised count of *CEBPA* expression (red dot indicates t(3;21) sample), in either normal karyotype AML (n=63) vs CBF fusion protein AML (t(8;21): n=7, t(3;21): n =1) from the TCGA project (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/# accessed 30th June 2016). Unpaired T-test used for comparison of gene expression values.

B) Normalised signal from gene expression microarray transformed by Log2 from Verhaak et al., 2009 (Verhaak et al., 2009) for *CEBPA* (red dot indicates t(3;21) sample). Normal karyotype AML (n=187) vs CBF fusion protein AML (t(8;21): n=35, t(3;21): n =1). Data downloaded from GEO, accession number GSE6891, accessed 30th July 2016.

C) Comparison of knockdown of CBF fusion protein in both types of AML identifies changes in biological processes common to both cell types. Hierarchical clustering of GO terms for up-regulated or down-regulated genes after either CBF fusion siRNA (siRUNX1-EVI1: RUNX1-EVI-1 siRNA; siRUNX1-ETO: RUNX1-ETO siRNA), as compared to control siRNA. Expression was measured after 10 days of treatment in SKH-1 cells or 4 days of treatment in Kasumi-1 cells. The threshold for differentially expressed genes was defined as 1.5 fold difference in expression between CBF fusion siRNA vs control siRNA treated cells.

D) RUNX1 and CBF fusion proteins in both t(3;21) and t(8;21) cells bind to the *CEBPA* +47Kb enhancer. UCSC browser screen at *CEBPA* of ChIP-seq of RUNX1, and either RUNX1-ETO in Kasumi-1 or RUNX1-EVI-1 in SKH-1.

3.6.3 RUNX1-EVI-1 knockdown results in a genome wide increase of C/EBPα binding.

Having characterised the changes in accessibility of cis-regulatory elements after RUNX1-EVI-1 knockdown, and up regulation of *CEBPA* expression following RUNX1-EVI-1 knockdown, we examined where C/EBPα was binding after RUNX1-EVI-1 knockdown. Identification of C/EBPα binding sites informs our understanding of the mechanism by which this transcription factor causes differentiation of the cells following RUNX1-EVI-1 knockdown. We therefore performed ChIP-seq for C/EBPα and also for RUNX1 and RUNX1-EVI-1 as controls, in SKH-1 cells after 10 days of either control or RUNX1-EVI-1 siRNA

transfection. We also performed a RUNX1 and RUNX1-EVI-1 ChIP-seq after 2 days of either control or RUNX1-EVI-1 siRNA transfection. The parameters of the ChIP-seq experiments (table 3-6) and annotation of binding sites are as shown below (table-3-6 and figure 3-17A). Examining specifically the RUNX1-EVI-1 bound regions, figure 3-17B depicts the average global profile of the different ChIP-seq experiments at these sites. As expected, RUNX1-EVI-1 binding decreased following RUNX1-EVI-1 knockdown. At the sites bound by RUNX1-EVI-1, RUNX1 binding was largely invariant. However, C/EBPα binding was greatly increased at sites previously bound by RUNX1-EVI-1.

Examples of the changes in C/EBPα binding after RUNX1-EVI-1 siRNA are depicted in figure 3-17C-D. Some of the target genes with increased binding by C/EBPα, following RUNX1-EVI-1 knockdown, are required for terminal myeloid differentiation, e.g. *MPO* and *CTSG*. These genes increased in expression, following RUNX1-EVI-1 knockdown. However, following RUNX1-EVI-1 knockdown, C/EBPα also increased binding to a number of genes which became repressed. These genes are required for stem cell function and are frequently deregulated in leukaemias, such as *GATA2, MEIS1*, and *MYCN* (Ye et al., 2013, Kawagoe et al., 2007). This suggests that increase in *CEBPA* expression after RUNX1-EVI-1 knockdown results in increased binding of C/EBPα at key genes that regulate haematopoietic differentiation and required for terminal myeloid function.

Table 3-6 ChIP seq in SKH-1 after 2 days (D2) or 10 days (D10) treatment of control (MM) or RUNX1-EVI-1 siRNA (KD)
High confidence peaks are ChIP-seq peaks that also coincide with DHSs.

ChIP-seq Dataset	Aligned reads (n)	Peaks (n)	High Confidence Peaks (n)
Anti-EVI1 D10 MM	20,350,942	12,842	7,252
Anti-EVI1 D10 KD	29,752295	2,014	1,208
Anti-RUNX1 D2 MM	12,556,575	14,510	12,142
Anti-RUNX1 D2 KD	7,123,119	14,165	10,983
Anti-RUNX1 D10 MM	19,147,420	15,185	10,346

Anti-RUNX1 D10 KD	20,959,942	15,592	9,944
Anti- C/EBPα D10 MM	23,509,134	9,346	7,016
Anti- C/EBPα D10 KD	29,564,903	11,808	8,639

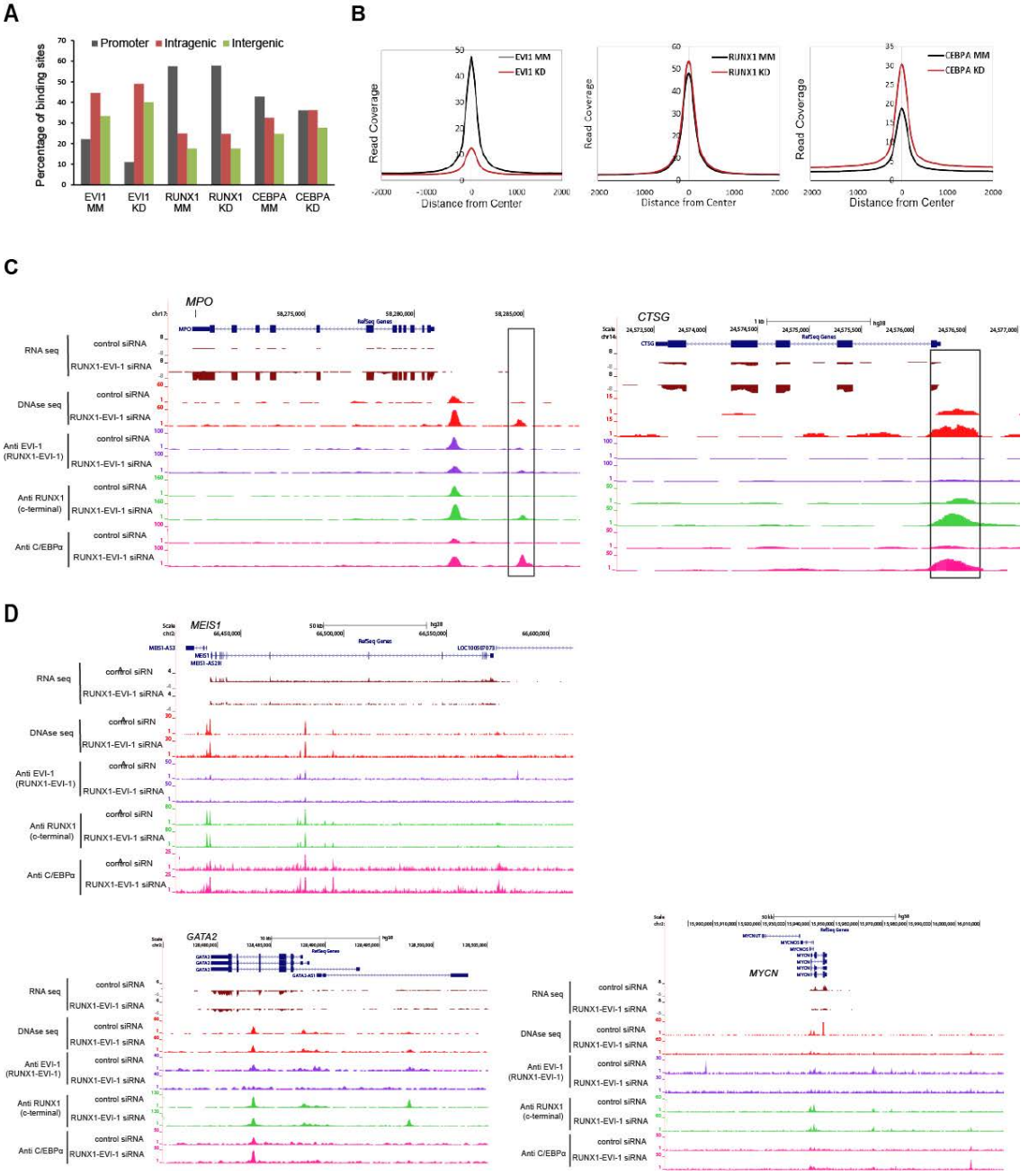


Figure 3-17 C/EBPα binding increases at both myeloid specific and stem cell associated genes after RUNX1-EVI-1 knockdown in SKH-1 cells

A) Genomic distribution of binding sites by promoter, intragenic and intergenic region for RUNX1, EVI-1 (RUNX1-EVI-1) and C/EBPα ChIP seq peaks in t(3;21) SKH-1 after 10 days of either control (MM) or RUNX1-EVI-1 (KD) siRNA.

B) Changes in binding of RUNX1 and C/EBPα at RUNX1-EVI-1 shared sites after RUNX1-EVI-1 knockdown in SKH-1 cells. RUNX1-EVI-1 DNA binding decreased, and C/EBPα binding increased at the same sites following RUNX1-EVI-1 knockdown Average profiles of EVI-1 (RUNX1-EVI-1), RUNX1 and

C/EBPα ChIP-seq centred on EVI-1 peaks in a 4 kb window in SKH-1 cells after10 days of treatment with either control (MM) or RUNX1-EVI-1 (KD) siRNA.

C-D) Changes in RUNX1-EVI-1, RUNX1 and C/EBPα binding in SKH-1 after RUNX1-EVI-1 knockdown. UCSC browser screen shots of the distribution of RNA-seq, DNase-seq and ChIP-seq aligned reads 10 days after either control siRNA or RUNX1-EVI-1 siRNA treatment c) *MPO* and *CTSG* and d) *MEIS1*, *MYCN* and *GATA2* loci. *MPO* and *CTSG* expression increased and this increase was accompanied by increases in DNase accessibility, RUNX1 and C/EBPα binding, following RUNX1-EVI-1. *MEIS1, GATA2* and *MYCN* expression decreased following RUNX1-EVI-1 knockdown which was accompanied by increases in C/EBPα binding at these loci.

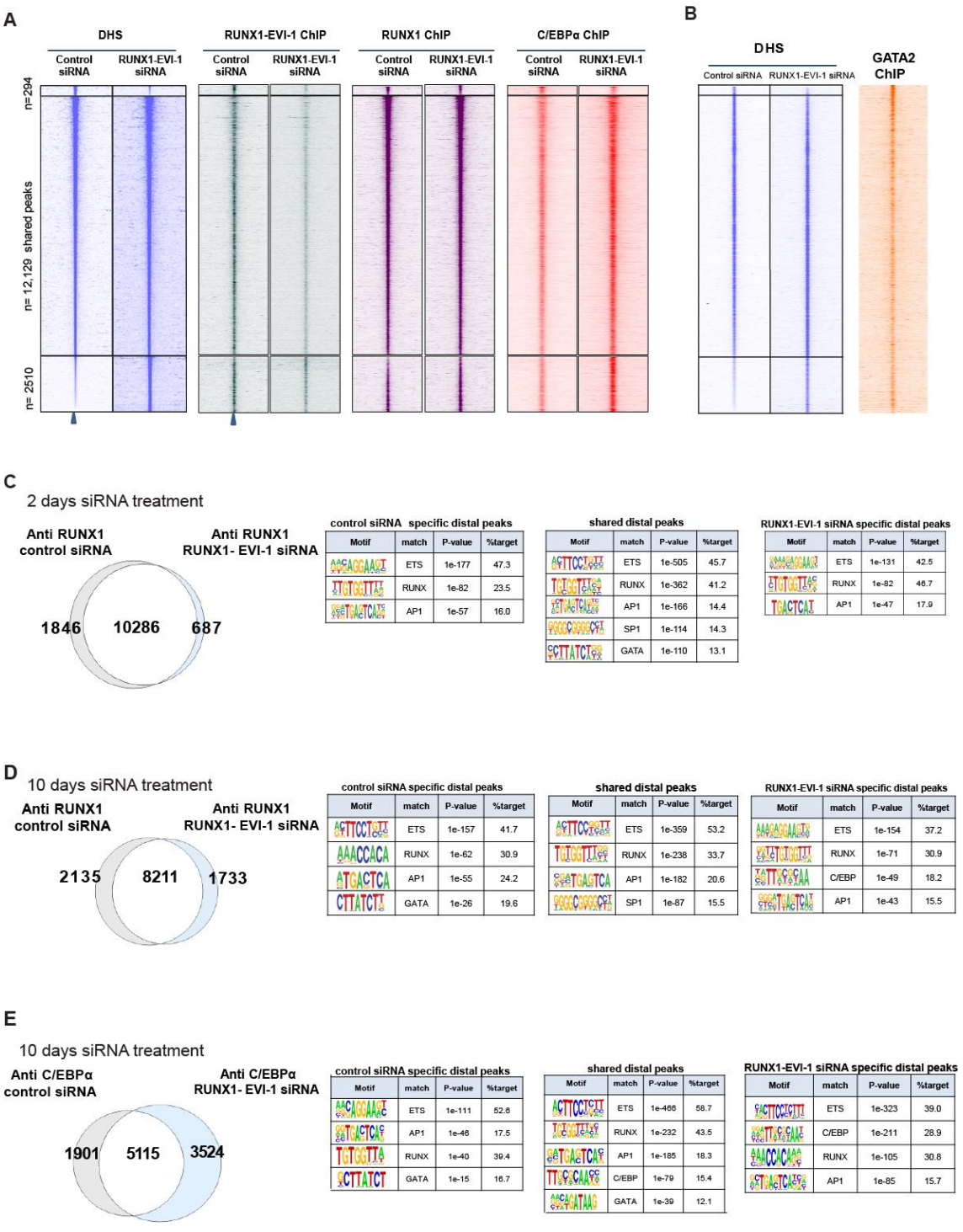
We noted that the increased expression of certain genes such as *MPO* and *CTSG* was accompanied by changes in DNaseI accessibility and increased binding of C/EBPα and RUNX1 (figure 3-17). However, it was unclear as to whether this was a genome wide phenomenon. To examine this relationship between the changes in DHSs and the differences in transcription factor binding following RUNX1-EVI-1 knockdown in more detail, we plotted a tag density heat map centred on the DHS peaks that were identified in SKH-1 cells transfected for 10 days with either control or RUNX1-EVI-1 siRNA,. Alongside the same genomic coordinates we plotted tag density heat maps from each of the ChIP-seq experiments (RUNX1-EVI-1, RUNX1 and C/EBPα in SKH-1 transfected for 10 days with either control or RUNX1-EVI-1 siRNA) (Figure 3-18A). This analysis shows that RUNX1-EVI-1 binding is globally depleted after knock-down. After RUNX1-EVI-1 siRNA transfection, RUNX1 binding was largely unchanged, except for an increase in binding in DHS which became more hypersensitive following knockdown of RUNX1-EVI-1. C/EBPα binding after RUNX1-EVI-1 knockdown was increased globally, but especially at the DHSs with increased hypersensitivity. Using the TF1 GATA2 ChIP-seq data we also noted that the DHSs with decreased hypersensitivity were enriched for GATA2 binding (figure 3-18B).

To understand whether the changes in transcription factor binding was caused by direct DNA binding or, indirectly through an intermediate transcription factor we analysed the underlying motifs of the transcription factor binding sites following RUNX1-EVI-1 knockdown. Analysis of enriched motifs within RUNX1

ChIP seq binding sites 2 days post siRNA treatment show little difference in motif enrichment with or without RUNX1-EVI1 knock-down (figure 3-18C). After 10 days of RUNX1-EVI-1 siRNA treatment, both RUNX1 and C/EBPα ChIP peaks showed a loss of GATA motifs. Following RUNX1-EVI-1 siRNA transfection, both RUNX1 and C/EBPα de novo binding sites were enriched for CEBP motifs (figure 3-18D-E). CEBP and GATA motifs were not found in RUNX1 binding sites present in both control and RUNX1-EVI-1 siRNA transfected cells.

This data suggests that RUNX1 and C/EBPα binding sites which are lost following RUNX1-EVI-1 knockdown, are bound in association with a GATA transcription factor, most likely GATA2. In contrast, the new RUNX1 binding sites appearing following RUNX1-EVI-1 knockdown are characterised by CEBP motifs. This is in keeping with the increase in CEBP and RUNX motif containing DHSs following RUNX1-EVI-1 knockdown.

The data taken together in this section suggests that the mechanism behind the changes in SKH-1 cells following RUNX1-EVI-1 knockdown is the up-regulation of *CEBPA* and its genome-wide binding to key loci causing the repression of stem cells genes and driving myeloid differentiation.



Motif	match	P-value	%target
	ETS	1e-177	47.3
	RUNX	1e-82	23.5
	AP1	1e-57	16.0

Motif	match	P-value	%target
	ETS	1e-505	45.7
	RUNX	1e-362	41.2
	AP1	1e-166	14.4
	SP1	1e-114	14.3
	GATA	1e-110	13.1

Motif	match	P-value	%target
	ETS	1e-131	42.5
	RUNX	1e-82	46.7
	AP1	1e-47	17.9

Motif	match	P-value	%target
	ETS	1e-157	41.7
AAACCACA	RUNX	1e-62	30.9
ATGACTCA	AP1	1e-55	24.2
CTTATCT	GATA	1e-26	19.6

Motif	match	P-value	%target
	ETS	1e-359	53.2
	RUNX	1e-238	33.7
	AP1	1e-182	20.6
	SP1	1e-87	15.5

Motif	match	P-value	%target
	ETS	1e-154	37.2
	RUNX	1e-71	30.9
	C/EBP	1e-49	18.2
	AP1	1e-43	15.5

Motif	match	P-value	%target
	ETS	1e-111	52.6
	AP1	1e-46	17.5
TGTGGTTA	RUNX	1e-40	39.4
GCTTATCT	GATA	1e-15	16.7

Motif	match	P-value	%target
	ETS	1e-466	58.7
	RUNX	1e-232	43.5
	AP1	1e-185	18.3
	C/EBP	1e-79	15.4
	GATA	1e-39	12.1

Motif	match	P-value	%target
	ETS	1e-323	39.0
	C/EBP	1e-211	28.9
	RUNX	1e-105	30.8
	AP1	1e-85	15.7

Figure 3-18 Global increase in C/EBPα binding after RUNX1-EVI-1 knockdown in SKH-1 cells

A) Heatmap showing read densities of ChIP-Seq peaks centred on DHS peaks in SKH-1 cells treated with control siRNA within +/-4Kb windows. Aligned to these coordinates are DHS peaks in SKH-1 cells treated with RUNX1-EVI-1

siRNA, followed by ChIP-seq of RUNX1-EVI-1, RUNX1 and C/EBPα in SKH-1 cells after either 10 days of either control or RUNX1-EVI-1 siRNA treatment. Increased DNaseI accessibility was seen in in 2510 sites following RUNX1-EVI-1 knockdown with increased binding of RUNX1 and C/EBPα at these sites.

B) Profiles of the DNase-seq tag density, in +/-4Kb windows centred on peaks, for SKH-1 treated for 10 days with either control or RUNX1-EVI-1 siRNA. Plotted to the right, GATA2 ChIP-seq tag density measured in the TF1 cell line (GSE73207), showing higher GATA2 binding at DHSs which decrease in accessibility following RUNX1-EVI-1 knockdown.

C-D) RUNX1-EVI-1 knockdown results in a loss of RUNX1 binding sites with GATA motifs and an increase in sites with enriched C/EBP motifs after 10 days of RUNX1-EVI-1 siRNA treatment. Left panel: Venn diagram showing peak overlaps from RUNX1 ChIP-seq after either C) 2 days or D) 10 days of control or RUNX1-EVI-1 siRNA transfection. The right panels show de novo motif analyses from distal peaks found after either siRNA treatment, or, specifically only after control siRNA or RUNX1-EVI-1 siRNA.

E) RUNX1-EVI-1 knockdown results in loss of C/EBPα binding sites with an enrichment of GATA motifs and an increase in sites with an enrichment of C/EBP motifs. Left panel: Venn diagram showing peak overlaps from C/EBPα ChIP-seq after either 10 days of control or RUNX1-EVI-1 siRNA. The right panels show de novo analyses from distal peaks found after either siRNA treatment, or, specifically only after control siRNA or RUNX1-EVI-1 siRNA.

3.7 The response to RUNX1-EVI-1 knockdown is blocked by dominant negative C/EBP (DNCEBP) peptide

3.7.1 Phenotypic changes of RUNX1-EVI-1 knockdown is inhibited by presence of DNCEBP

To investigate whether C/EBPα DNA binding was required for the downstream effects of RUNX1-EVI-1 knockdown, we generated SKH-1 cells expressing a dominant negative CEBP peptide (DNCEBP). This DNCEBP peptide has been shown to dimerise with all CEBP transcription factors and prevent binding to DNA (Krylov et al., 1995). To this end, we transduced SKH-1 cells with lentiviruses carrying either an empty or a vector expressing a FLAG epitope tagged DNCEBP peptide. Expression of DNCEBP peptide was confirmed by Western blotting (figure3-19A) using an antibody against the FLAG tag. We then serially transfected the transduced SKH-1 cells with either control or

RUNX1-EVI-1 siRNA over a 14 day period (figure 3-19B). Both empty vector and DNCEBP vector transduced SKH-1 cells were treated in parallel. Knockdown of RUNX1-EVI-1 in each cell line was confirmed by Western blot and RT-qPCR (figure 3-19A and B).

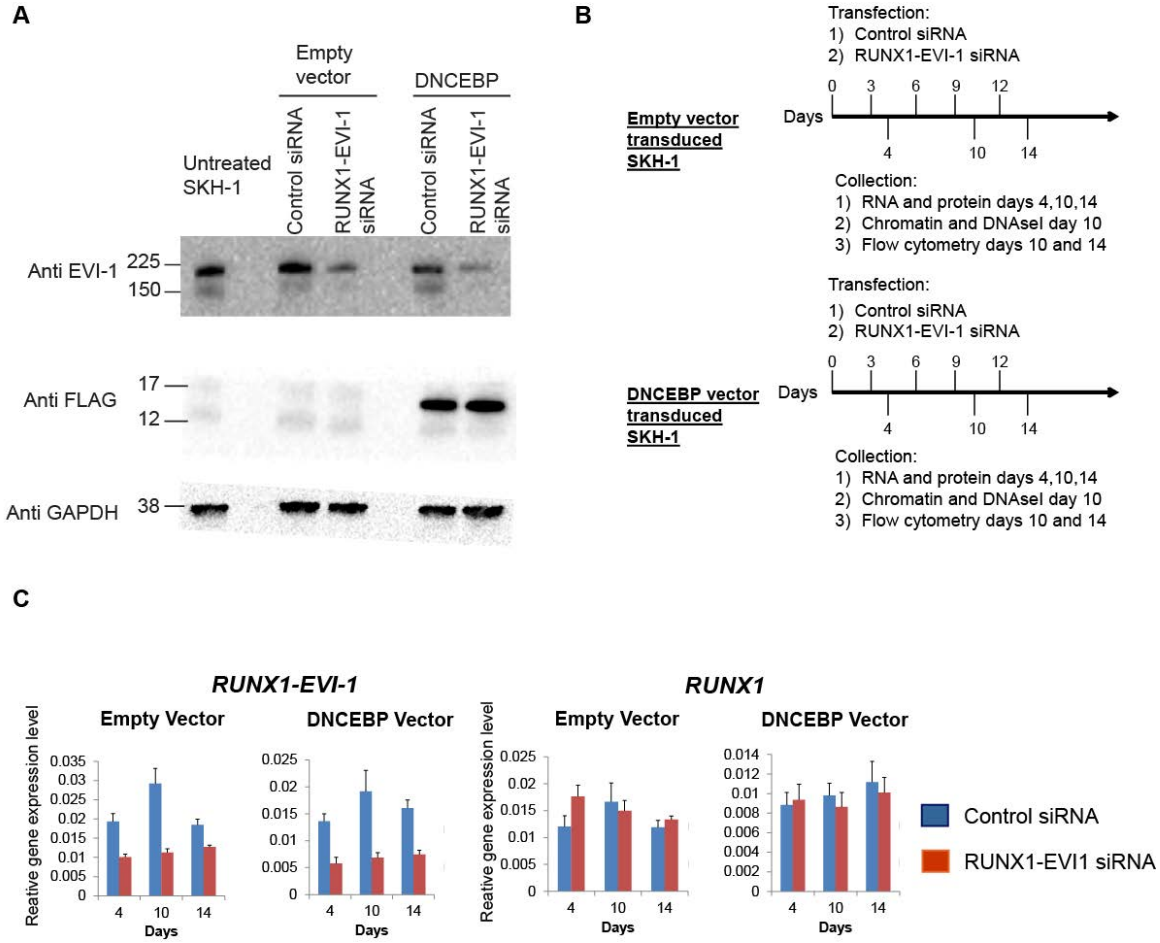


Figure 3-19 The down-stream effects of RUNX1-EVI-1 knock-down are blocked in SKH-1 cells transduced with a DNCEBP peptide

A) Western blot with whole cell lysates from untreated, empty or DNCEBP vector transduced SKH-1 cells, transfected with either control or RUNX1-EVI-1 siRNA after 14 days. The blot was probed with EVI-1, FLAG or GAPDH antibodies as indicated. Knock-down of RUNX1-EVI-1 was performed in both empty vector and DNCEBP vector expressing SKH-1. DNCEBP peptide expression was confirmed by probing with an anti-FLAG antibody (middle panel).

B) Experimental scheme for performing RUNX1-EVI-1 knockdown in either empty vector or DNCEBP vector transduced SKH-1 cells with transfection with

either control siRNA or RUNX1-EVI-1 siRNA being performed on days 0, 3, 6, 9 or 12. Cells were harvested to collect for RNA, chromatin, DNaseI or flow cytometry on indicated days.

C) Analysis of mRNA levels relative to *GAPDH* after control or RUNX1-EVI-1 siRNA transfection (4, 10 or 14 days of treatment). *RUNX1-EVI-1* or *RUNX1* mRNA levels in either empty vector or DNCEBP vector transduced SKH-1 were measured at the indicated days, showing a reduction of *RUNX1-EVI-1* mRNA in both empty and DNCEBP vector transduced cells following RUNX1-EVI-1 siRNA transfection.

SKH-1 cells that were transduced with an empty vector showed a decrease in CD34 expression and a decline in growth rate when serially transfected with RUNX1-EVI-1 siRNA as compared to control cells. In contrast, SKH-1 cells transduced with a vector expressing the DNCEBP peptide displayed neither a change in surface levels of CD34 nor a decline in growth (figure 3-20A-C). We measured the expression of the key stem cell genes *HOXA9, GATA2* and *MEIS1* which all decreased in empty vector transduced SKH-1 following knockdown of RUNX1-EVI-1. However, in DNCEBP transduced SKH-1 there was no change in *HOXA9* expression following RUNX1-EVI-1 knockdown. Expression of *GATA2* and *MEIS1* still declined, but this remained expressed at a higher level (figure 3-20D). *CTSG, MPO*, and *CSF1R*, all of which are markers of terminal myeloid differentiation, decreased as expected in empty vector transduced SKH-1 following RUNX1-EVI-1 sRNA transfection, whilst in SKH-1 expressing DNCEBP, RUNX1-EVI-1 knock-down had no effect on the expression of these genes (figure 3-20E). It is therefore clear that the phenotypic changes following RUNX1-EVI-1 (figure 3-9) require C/EBPα DNA binding. However, despite up regulation of these genes required for myeloid differentiation, neither CD14 nor CD11b expression changed in empty vector transduced SKH-1 following RUNX1-EVI-1 siRNA transfection, suggesting the differentiation process remained incomplete.

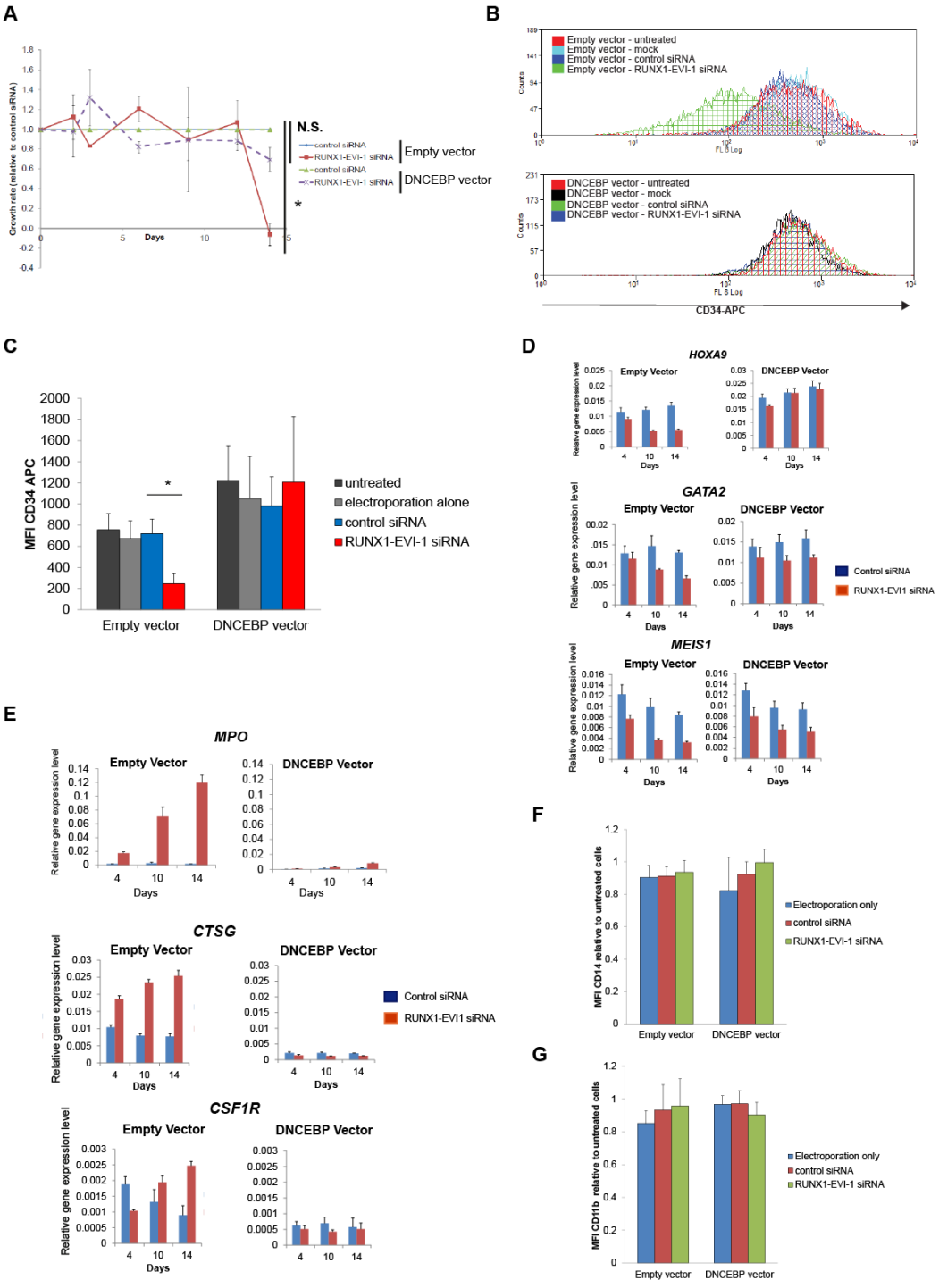


Figure 3-20 DNCEBP expression blocks the phenotypic changes seen in SKH-1 cells following RUNX1-EVI-1 knockdown

A) Reduction in cell growth by RUNX1-EVI-1 siRNA is abrogated by DNCEBP peptide. Growth rate of empty (solid lines) or DNCEBP (dashed lines) vector transduced SKH-1, transfected with RUNX1-EVI-1 siRNA, relative to control siRNA treatment. The graph shows the mean and SEM of 3 independent

experiments. * denotes p<0.05 by unpaired t-test between RUNX1-EVI-1 siRNA transfected empty and DNCEBP vector transduced SKH-1, and paired t-test between control and RUNX1-EVI-1 siRNA transfected empty vector transduced SKH-1 cells, after 14 days of siRNA treatment.

B-C) DNCEBP prevents downregulation of CD34 by RUNX1-EVI-1 siRNA transfection. Flow cytometry of empty and DNCEBP vector transduced SKH-1 cells, untreated or after 14 days of either mock, control or RUNX1-EVI-1 siRNA treatment. B) Representative overlay of histogram with CD34-APC staining, C) The graph shows the MFI (median), mean of 3 independent experiments with error bars representing SEM. *denotes p<0.05 by paired t-test.

D) DNCEBP prevents downregulation of *HOXA9* and abrogates down-regulation of *GATA2* and *MEIS1* expression. mRNA levels by RT-qPCR relative to *GAPDH* after RUNX1-EVI-1 siRNA transfection (4, 10 or 14 days of treatment). *HOXA9* expression decreased in empty vector transduced SKH-1 transfected with RUNX1-EVI-1 siRNA as compared to control siRNA, but this decrease is abrogated in DNCEBP vector transduced SKH-1. *GATA2* and *MEIS1* follow a similar pattern but the impact of DNCEBP is reduced. The graph shows the mean and SEM of 3 independent experiments.

E) DNCEBP prevents up-regulation of *MPO*, *CSF1R* and *CTSG* following RUNX1-EVI-1 knockdown. mRNA levels by RT-qPCR relative to *GAPDH* after control RUNX1-EVI-1 siRNA transfection (4, 10 or 14 days of treatment). *MPO*, *CSF1R* and *CTSG* increases after transfection with RUNX1-EVI-1 siRNA, as compared to control siRNA, in empty vector transduced cells but not in DNCEBP vector transduced cells. The graph shows the average and SEM of 3 independent experiments.

F-G) Myelopoiesis following RUNX1-EVI-1 knockdown is incomplete. Flow cytometric analysis of either empty or DNCEPB vector transduced SKH-1 cells untreated, or after 14 days of either electroporation alone, or after control or RUNX1-EVI-1 siRNA transfection. MFI (median) of either F) CD14 or G) CD11b staining, relative to non-electroporated cells after 14 days of treatment. The graph shows the mean and SEM of 3 independent experiments.

3.7.2 DNCEBP expression alters regulation of gene expression by C/EBPα

We demonstrated in figure 3-17 and 3-18 that DHSs that increase in nuclease accessibility were associated with increases in C/EBPα and RUNX1 binding. However, it was unclear whether this increase in DHS hypersensitivity could be seen in the absence of CEBPα, and whether CEBPα binding was required for RUNX1 binding. We chose examples of several loci where DHS became more

accessible (*MPO* enhancer, *CTSG* promoter (figure 3-17C); *TREM1* promoter, *SIGLEC1* enhancer (figure3-21A)). These sites are associated with genes expressed in mature myeloid cells (Borregaard et al., 2007, Barral et al., 2010, Klesney-Tait et al., 2013). C/EBPα bound to these sites after RUNX1-EVI-1 knockdown in empty vector transduced SKH-1, but in the presence of DNCEBP binding was blocked. Similarly, RUNX1 bound to these sites in empty vector transduced SKH-1 cells following RUNX1-EVI-1 knockdown, but in the absence of C/EBPα binding in DNCEBP expressing SKH-1, RUNX1 was unable to bind to these sites (figure 3-21B). Finally, following RUNX1-EVI-1 knockdown, chromatin remains inaccessible at these sites in DNCEBP vector expressing SKH-1 cells, as compared to control cells (figure 3-21C). This result suggests that following RUNX1-EVI-1 knock-down and *CEBPA* up-regulation, C/EBPα DNA binding cooperates with RUNX1 to increase chromatin accessibility to up-regulate genes required for terminal myeloid function.

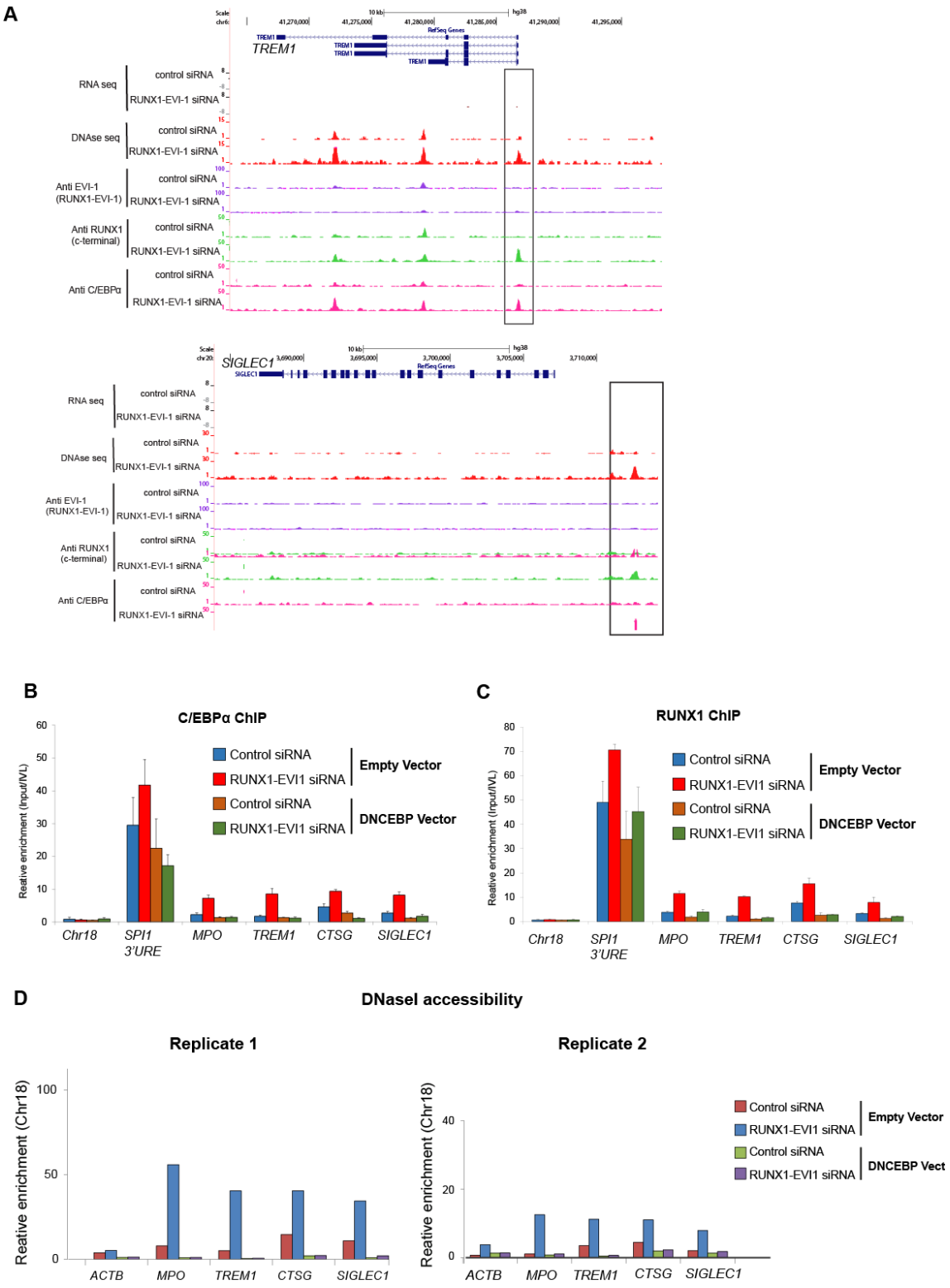


Figure 3-21 C/EBPα DNA binding is required to recruit RUNX1 and open previously inaccessible chromatin

A) UCSC browser screen-shot showing changes in RUNX1-EVI-1, RUNX1 and C/EBPα binding after RUNX1-EVI-1 knockdown at *TREM1* and *SIGLEC1* loci. Increased chromatin accessibility following RUNX1-EVI-1 knockdown is associated with increased C/EBPα and RUNX1 DNA binding at these loci. Boxes designate sites used for amplicons in ChIP and DNaseI experiments described in B) and C). RNA-seq, DNase seq and ChIP seq were performed 10 days after either control siRNA or RUNX1-EVI-1 siRNA treatment.

B-C) Increased C/EBPα, and RUNX1 binding after RUNX1-EVI-1 knockdown is prevented by DNCEBP expression. B) C/EBPα or C) RUNX1 ChIP -qPCR using amplicons corresponding to the *MPO* and *SIGLEC1* enhancers and the *TREM1* and *CTSG* promoters. The position of amplicons is highlighted by boxes in the UCSC browser screenshots from figure 3-17C, and figure 3-21A. An amplicon covering the *SPI1 3' URE* was used as positive control and chromosome 18 as negative control. qPCR enrichment relative to input and *IVL*. Chromatin was harvested from empty or DNCEBP vector transduced cells 10 days after either control or RUNX1-EVI-1 siRNA transfection. The graph shows the mean of 3 independent experiments, with error bars representing SEM.

c) Increased DNaseI hypersensitivity after RUNX1-EVI-1 knockdown is prevented by DNCEBP. DNaseI and qPCR validation at *MPO* and *SIGLEC1* enhancer, *TREM1* and *CTSG* promoter. Amplicons covering the *ACTB* gene body were used as as negative control. qPCR enrichment was calculated relative to chromosome 18. 2 independent experiments involving 10 days of siRNA transfection are shown.

3.8 Overexpression of an inducible version of C/EBPα phenocopies RUNX1-EVI-1 knockdown

Having demonstrated that C/EBPα DNA binding capacity is critical for the downstream effects of RUNX1-EVI-1 knockdown, we next investigated whether over-expression of C/EBPα was sufficient to replicate the effects of RUNX1-EVI-1 knockdown. To accomplish this we used a retrovirus encoding a *CEBPA-ER* transgene fused to an IRES-GFP sequence. C/EBPα –ER is a fusion of C/EBPα with the estrogen receptor ligand binding domain which can translocate into the nucleus when bound to 17β-Estradiol (Bussmann et al., 2009, Umek et al., 1991). Successfully transduced SKH-1 were identifiable by the presence of GFP and separated by FACS. A Western blot with lysates from C/EBPα-ER expressing SKH-1 cells show that they express the correct protein size (76kDa) (figure 3-22A). RT-PCR using C/EBPα-ER specific amplicons showed that

mRNA expression is specific in SKH-1 transduced with virus, as compared to mock treated cells (figure 3-22B). The levels of C/EBPα-ER mRNA appeared to be decreased in the cells treated with 10nM 17β-Estradiol because the GFP positive population of cells declined following treatment as described previously in the methods (figure 2-3). However, after induction of C/EBPα activity, the endogenous CEBPA gene showed an upregulation as a result of an "auto-regulation" of this gene which has been seen before (Timchenko et al., 1995) (figure 3-22B).

C/EBPα-ER SKH-1 cells were exposed to 10nM 17β-estradiol for 48 hrs which resulted in an increase of apoptotic cells (figure 3—22 C-D) and reduction in growth rate (figure 3-22E), when compared to ethanol treatment, or to mock treated SKH-1 exposed to the same treatments. This increase of apoptosis was similar to the effects of RUNX1-EVI-1 knockdown seen in figure 3-10.

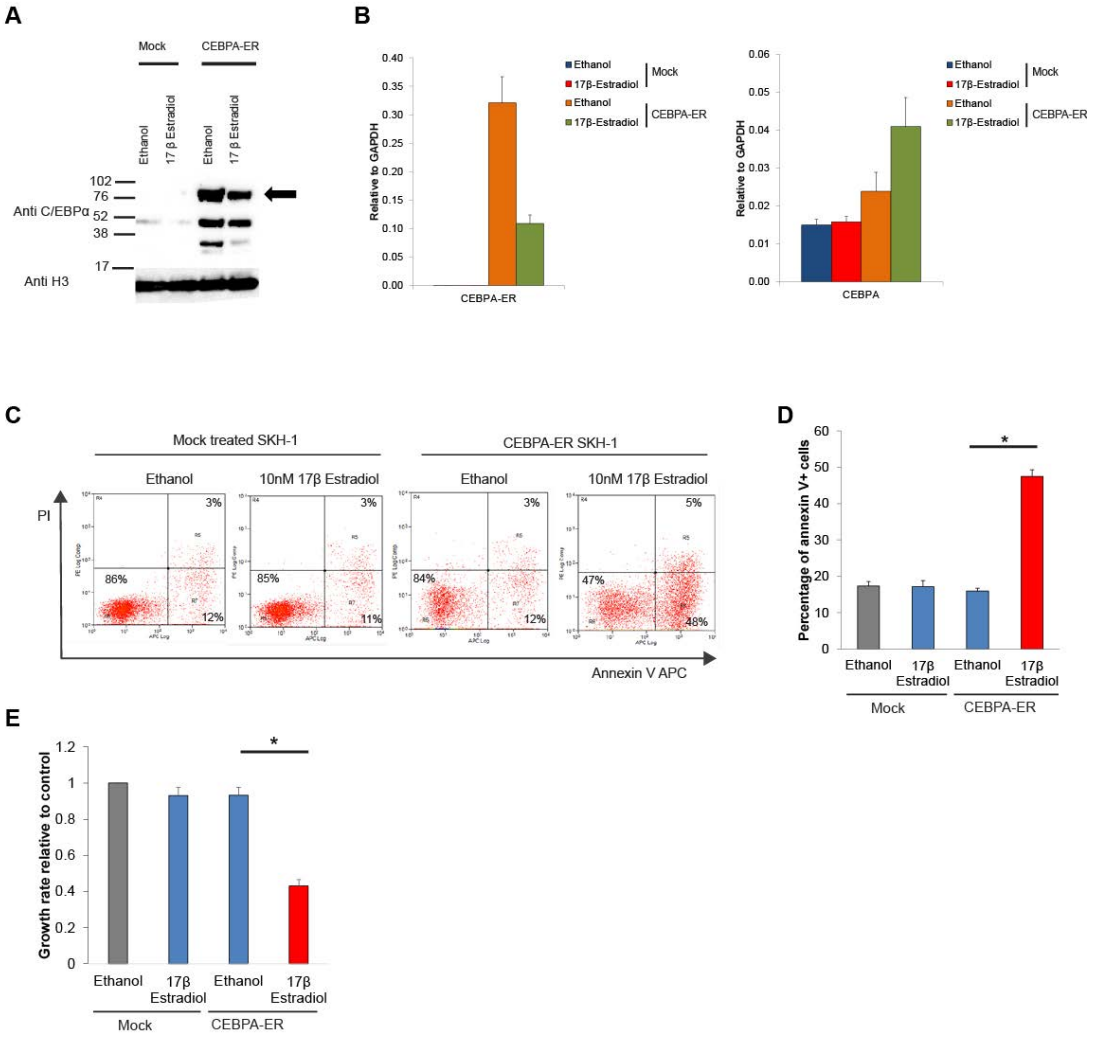


Figure 3-22 Induction of C/EBPα-ER by phenocopies RUNX1-EVI-1 knockdown

A-B) CEBPA-ER transduction in SKH-1 cells results in expression of a fusion C/EBPα –ER protein.

A) Western blot with nuclear extracts from mock treated or CEBPA-ER transduced SKH-1 cells treated with either vehicle (ethanol) or 10nM 17β-Estradiol. The membrane was probed with C/EBPα and H3 antibodies as indicated. The black arrow indicates the size of band corresponding to the C/EBPα-ER fusion protein.

B) Endogenous *CEBPA* and *CEBPA-ER* mRNA levels as measured by RT-qPCR relative to *GAPDH* of mock treated or CEBPA-ER transduced SKH-1 cells treated with either vehicle (ethanol) or 10nM 17β-Estradiol. The graph shows the mean of 3 independent experiments with SEM.

C-E) Induction of C/EBPα activity increases apoptosis in CEBPA-ER SKH-1 cells. Annexin-V APC and PI staining in mock treated or CEBPA-ER transduced SKH-1 cells, treated with either vehicle (ethanol) or 10nM 17 β – Estradiol followed by flow cytometry. C) Representative plots are shown. D) Relative frequencies of unstained, Annexin-V positive cells. E) Growth rate relative to ethanol treated mock transduced SKH-1 after 2 days of treatment of either ethanol or 10nM 17 β -Estradiol. The Graph shows the mean and SEM of 3-5 independent experiments. * denotes $p<0.05$ by paired t-test.

RUNX1-EVI-1 knockdown resulted in a down-regulation of CD34 and genes associated with stem cell function, alongside up-regulation of genes associated with myelopoiesis (figure 3-9). Activation of C/EBPα-ER phenocopied RUNX1-EVI-1 knockdown and led to the down-regulation of CD34 and genes associated with stem cell function (figure 3-23 A-C), despite the continued expression of RUNX1-EVI-1 (figure 3-23 C). In parallel, genes associated with terminal myeloid differentiation were up-regulated following induction of C/EBPα-ER (figure 3-23C). This was accompanied by an up-regulation of CD11b (figure 3-23 D-E). Colony forming capacity was also reduced following C/EBPα-ER induction, in terms of both total numbers (figure 3-23 F) as well as morphologically in terms of size (figure 3-23 G).

Therefore, overexpression of C/EBPα alone is sufficient to drive the changes seen in SKH-1 cells following RUNX1-EVI-1 knockdown. This evidence along with the data provided from the DNCEBP expressing SKH-1 cell line (Chapter 3.7), suggests that inhibition of C/EBPα by RUNX1-EVI-1 is the critical node by which t(3;21) cells maintain their leukaemic identity. Our data also show that both RUNX1-EVI-1 and RUNX1-ETO share this C/EBPα dysfunction (Chapter 3.6.2).

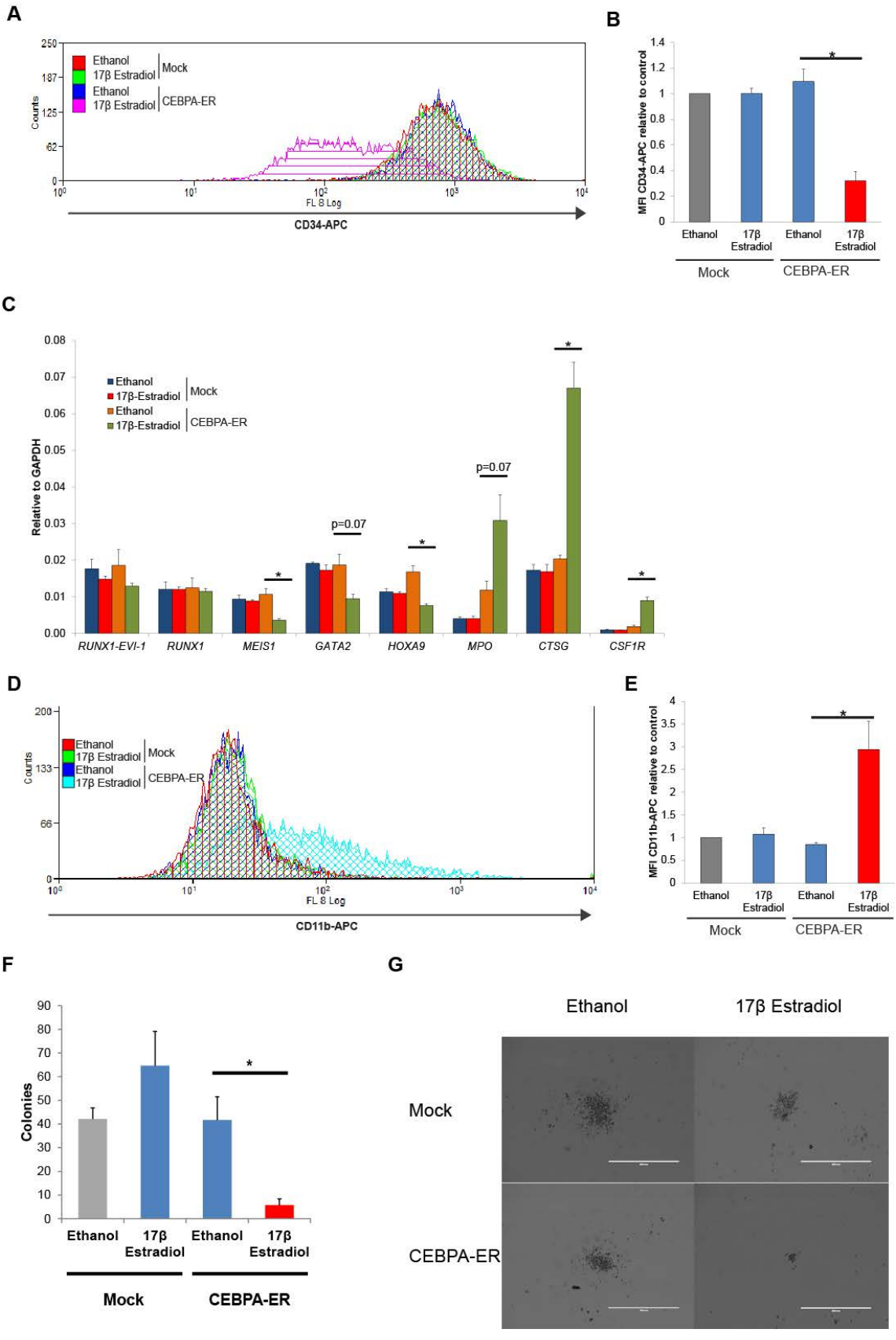

A
Ethanol
17β Estradiol
Mock
Ethanol
17β Estradiol
CEBPA-ER
Counts
FL 8 Log
CD34-APC
B
MFI CD34-APC relative to control
Ethanol
17β Estradiol
Ethanol
17β Estradiol
Mock
CEBPA-ER
C
Ethanol
17β-Estradiol
Mock
Ethanol
17β-Estradiol
CEBPA-ER
Relative to GAPDH
p=0.07
p=0.07
RUNX1-EVI-1
RUNX1
MEIS1
GATA2
HOXA9
MPO
CTSG
CSF1R
D
Ethanol
17β Estradiol
Mock
Ethanol
17β Estradiol
CEBPA-ER
Counts
FL 8 Log
CD11b-APC
E
MFI CD11b-APC relative to control
Ethanol
17β Estradiol
Ethanol
17β Estradiol
Mock
CEBPA-ER
F
Colonies
Ethanol
17β Estradiol
Ethanol
17β Estradiol
Mock
CEBPA-ER
G
Ethanol
17β Estradiol
Mock
CEBPA-ER

Figure 3-23 C/EBPα-ER activation overrides the effects of RUNX1-EVI-1 and leads to partial differentiation of SKH-1 cells

A-B) Induction of C/EBPα activity in CEBPA-ER SKH-1 cells results in the down-regulation of CD34. Flow cytometry of mock or CEBPA-ER transduced SKH-1 cells treated with either vehicle (ethanol) or 10nM 17β-Estradiol for 48 hours. A) Representative overlay of histogram of cells stained with CD34-APC. B) MFI (median) of CD34-APC staining. MFI relative to ethanol treated, mock transduced SKH-1. The graph shows the mean of 3 independent experiments with SEM. *denotes $p<0.05$ by paired t-test.

C) Changes in gene expression of key target genes after induction of C/EBPα-ER activity mirrors the changes seen after RUNX1-EVI-1 knockdown. mRNA levels relative to GAPDH of the indicated genes were measured in mock treated or CEBPA-ER transduced SKH-1 cells treated with either vehicle (ethanol) or 10nM 17β-Estradiol. The graph shows the mean of 3 independent experiments with SEM. * denotes $p<0.05$ by paired t-test.

D-E) Induction of C/EBPα activity in CEBPA-ER SKH-1 cells results in the up-regulation of CD11b. Flow cytometry of mock or CEBPA-ER transduced SKH-1 cells were treated with either vehicle (ethanol) or 10nM 17β-Estradiol. A) Representative overlay of flow cytometry histograms of cells stained with CD11b-APC. B) MFI (median) of CD11b-APC staining relative to ethanol treated, mock transduced SKH-1. The graph shows the mean of 3 independent experiments with SEM. *denotes $p<0.05$ by paired t-test.

F-G) Induction of CEBPA-ER results in reduced colony forming capacity. SKH-1 cells were either mock treated or transduced with CEBPA-ER virus. Mock treated or CEBPA-ER transduced SKH-1 were treated for 48 hours with either ethanol (vehicle control) or 17β Estradiol, prior to plating on methylcellulose culture. F) Colonies of over 20 cells were counted between 7-11 days after plating. The graph shows the mean of 5 experiments, with error bars representing S.E.M. * signifies $p<0.05$ by paired t-test. G) Representative photographs of colonies are shown. The scale bar represents 400 μM.

Chapter 4. Discussion

4.1 Epigenetic landscape differ between t(3;21) and t(8;21) leukaemias driving the use of different regulators of self-renewal

In this work, we demonstrated that t(3;21) leukaemia and t(8;21) leukaemia differ in their gene expression profiles and DHS landscape. The samples cluster according to gene expression profile in a similar pattern to the clustering based on the pattern of DHSs (figure3-2 and 3-3). This is not surprising, as these sites represent active cis-regulatory elements that control gene expression. This is consistent with previous studies showing that AML can be classified according to gene expression and methylation profiles that segregate samples according to their mutational status, suggesting that AML is a disease of epigenetic dysfunction (Figueroa et al., 2010b, Valk et al., 2004). This has been confirmed by the discovery from targeted and whole genome sequencing studies that mutations in AML primarily affect transcription factors and epigenetic regulators (Papaemmanuil et al., 2016, Cancer Genome Atlas Research, 2013).

In our study, we show examples of genes which are both highly expressed in t(8;21) leukaemia (*ROBO1* and *CACNA2D2*) as compared to t(3;21) leukaemia, and also genes which are poorly expressed in t(8;21) leukaemia as compared to t(3;21) leukaemia (*HOXA9* and *MEIS1*) (figure 3-1). t(3;21) leukaemia has not been studied before in this manner, but other studies have shown that t(8;21) AML segregate independently from other types of leukaemia (Valk et al., 2004). *ROBO1* and *CACNA2D2* are examples of genes known to be upregulated in t(8;21) leukaemia in comparison to other forms of AML. In contrast, *HOXA9* and *MEIS1* have previously been shown to be downregulated in t(8;21) leukaemia, as compared to other forms of AML (Lavallée et al., 2016, Ross et al., 2004, Valk et al., 2004, Yagi et al., 2003).

t(3;21) leukaemias have a poorer prognosis than t(8;21) leukaemia, with low remission rates of 43% and event free survival of only 14% (Lugthart et al., 2010). *HOXA9* and *MEIS1* are both expressed in t(3;21) leukaemia but not in t(8;21) leukaemia. *HOXA9* expression is associated with poor prognosis in AML (Golub et al., 1999, Andreeff et al., 2008) and its expression is linked to a number of mutational subtypes including *MLL* and *NUP98* translocations (Collins and Hess, 2015). *MEIS1* expression is also associated with poor prognosis AML as part of gene expression pattern (Eppert et al., 2011) seen in HSCs and LSCs. *HOXA9* and *MEIS1* are often co-expressed in AML (Lawrence et al., 1999) and *Hoxa9* requires the co-expression of *Meis1* to transform murine bone marrow progenitor cells (Kroon et al., 1998). This cooperativity can be explained by the identification of large number of cis-regulatory elements that are co-bound by both HOXA9 and MEIS1 (Huang et al., 2012). In turn, there is an association between EVI-1 (partner of RUNX1 in RUNX1-EVI-1) and HOXA9 and MEIS1. *HOXA9* and *MEIS1* are important downstream targets of MLL-fusion proteins and EVI1 have been shown to be important in maintaining the leukaemic growth of MLL-AF9 and MLL-ENL transformed cells (Bindels et al., 2012, Goyama et al., 2008). Cooperation between HOXA9 and MEIS1 with EVI-1 is suggested by the frequency of retroviral insertion sites at EVI-1, in retrovirus mediated over-expression of HOXA9 and MEIS1 (Jin et al., 2007).

In contrast, t(8;21) leukaemia have been shown by independent cooperative groups to have a significantly improved prognosis compared to other leukaemias (Grimwade et al., 2010, Byrd et al., 2002, Slovak et al., 2000). Remission rates have been reported as over 90% with 10 year overall survival at 61% (Grimwade et al., 2010). Interestingly, both *ROBO1* and *CACNA2D2* are associated with potential tumour suppressor function, as candidate genes in regions frequently deleted in solid tumours (Hesson et al., 2007) . ROBO1 is a receptor that binds the SLIT2 ligand and is important in neuronal development by ensuring correct axonal guidance. In MDS, loss of function mutations of *ROBO1* and *ROBO2* have been associated with progression to AML, and overexpression of ROBO1 in leukaemic cell lines led to increased apoptosis (Xu

et al., 2015). *CACNA2D2* encodes a voltage gated calcium channel, which when mutated in mice leads to the development of an ataxic gait and epilepsy (Barclay et al., 2001). No studies on *CACNA2D2* have been performed in AML, however, functional studies have shown both tumour suppressor activity in lung cancer (Carboni et al., 2003) and oncogenic properties in prostate cancer (Warnier et al., 2015). Overall, *CACNA2D2* is thought to be a tumour suppressor gene, as it is part of a 120kb region recurrently deleted in a variety of solid tumours (Hesson et al., 2007).

Therefore in this work we provide evidence of differentially expressed genes, that helps explain the difference in prognoses between the CBF leukaemia. Furthermore, we describe the mechanisms that maintain this gene expression profile by the use of DNase-seq. The identification of a pattern of cis-regulatory elements that distinguishes the t(8;21) from the t(3;21) leukaemia is further proof that this technology can be used for "class discovery" (Golub et al., 1999), enabling further refinement of AML subtyping. Here we demonstrate the advantage of DNase-seq over only using gene expression profiling as we are able to identify the cis-regulatory elements that define each leukaemia. For example, DHSs which are unique to t(3;21) leukaemia as compared to t(8;21) leukaemia frequently contain the GATA motif. In contrast, DHSs found in t(8;21) leukaemia as compared to t(3;21) leukaemia often contain the E-Box motif (figure 3-2D and E). This is directly related to the different properties of each CBF fusion protein (Section 4.3), thereby giving insight into the mechanism of leukaemogenesis in each type of CBF leukaemia.

The t(3;21) SKH-1 cell line appears to be an extremely good model for studying the epigenetic effects of t(3;21) leukaemia as the DHSs in the cell line and patient samples overlap substantially (figure 3-2B). However, in terms of gene expression, the t(3;21) SKH-1 cell line shows some distinct differences with the t(3;21) AML patient samples, in both a PCA analysis and clustering based on gene expression signal (figure 3-3 B and C). This may be due to intrinsic differences between autonomously growing cell lines as compared to primary patient samples, but also to the differences in cooperative mutations in these

samples. The SKH-1 cell line is derived from a patient with CML, who developed a myeloid blast crisis (Mitani et al., 1994) and therefore also expresses the translocation t(9;22) which produces the potent fusion protein BCR-ABL.

4.1.1 Distinct and common mechanisms for self-renewal between t(3;21) and t(8;21) leukaemia

In both RNA-seq and DNase-seq analyses, t(3;21) cluster with normal CD34+ progenitor cells, demonstrating that both cell types share a significant gene expression and cis-regulatory program. Although *CD34* expression is found in both HSCs and early progenitors, after knockdown of RUNX1-EVI-1in SKH-1 cells, a HSC and LSC gene expression gene sets are lost (figure 3-11D), thereby implicating this immature stem cell expression program in the pathogenesis of t(3;21) leukaemia. This may reflect shared "stemness" properties such as quiescence, niche dependence and self-renewal properties used in both HSCs and t(3;21) leukaemia. This is consistent with previous studies which have shown that LSCs and HSCs share a set of genes that likely represent these functions (Eppert et al., 2011). Previous literature may suggest that this is due to a cell of origin transformation being a normal HSC (Lapidot et al., 1994, Bonnet and Dick, 1997). Contemporary views suggest that committed progenitors may regain aspects of stem cell properties through transformation into AML (Goardon et al., 2011, Quek et al., 2016, Jamieson et al., 2004). However, a recent study suggests that HSCs transformed by MLL-AF9 differ from GMP transformed by the same fusion gene, by the higher expression of *EVI-1* and *ERG*, and this results in a more aggressive leukaemia (Stavropoulou et al., 2016). In turn, this is consistent with previous studies showing that EVI-1 overexpression is a poor prognostic factor in AML with MLL translocations (Groschel et al., 2013). Our GSEA demonstrating the loss of both HSC and LSC gene programs following knockdown of RUNX1-EVI-1, directly links RUNX1-EVI-1 to the maintenance of this gene expression profile, suggesting this oncogene has the potential to endow or maintain these "stem-like" properties.

A key member of this gene expression program found in both LSCs and HSCs (Eppert et al., 2011), and one of five genes expressed in both CD34+ and – LSCs (Quek et al., 2016) is *MEIS1,* which also play an important role in normal HSC biology. *Meis1* knockout mice die from haemorrhage due to a lack of megakaryocytes. Although myeloerythroid lineages develop, there is a reduction in colony forming potential and fetal liver cells from *Meis1* knockout mice have deficiencies in competitive repopulation assays in transplantation into lethally irradiated mice (Hisa et al., 2004). We demonstrate here that RUNX1-EVI-1 binds to the promoter of *MEIS1* and upon knockdown of RUNX1-EVI-1; *MEIS1* expression is down-regulated (figure 3-17D).

Our finding that RUNX1-EVI-1 is directly involved in regulating this stem cell program is entirely consistent with our knowledge of the fusion partner of RUNX1. EVI-1, also known as *MECOM*, is expressed in both HSCs and LSCs (Eppert et al., 2011). Its role in AML is well characterised. Over-expression of *EVI-1* is an independent poor prognostic factor in AML as demonstrated by a large cooperative group (Groschel et al., 2010). When retrovirally expressed in murine bone marrow it causes a myelodysplastic syndrome (Buonamici et al., 2004). It is also a key regulator of HSC function; *EVI-1* homozygous deficient HSCs have deficient self-renewal properties as shown by defective repopulating ability (Goyama et al., 2008, Kataoka et al., 2011).

Although t(8;21) leukaemia cluster separately, both RUNX1-ETO and RUNX1-EVI-1 target *ERG* (Ptasinska et al., 2014) (figure 3-12C). In both RUNX1-ETO and RUNX1-EVI-1 knockdown, expression of *ERG* is decreased. *ERG* is a member of the ETS family of transcription factors and is another part of both HSC and LSC gene expression program (Eppert et al., 2011). *ERG* overexpression in AML with normal karyotype is associated with a poor prognosis (Metzeler et al., 2009). ERG function is vital for the emergence of definitive HSCs and the ability of adult HSCs to repopulate lethally irradiated recipient mice (Loughran et al., 2008). Ectopic over expression of *ERG* in human CD34+ cord blood cells results in an expansion of the progenitor pool (Tursky et al., 2015).

WT1 is another target of both RUNX1-ETO and RUNX1-EVI-1 that is highly expressed in primitive CD34+CD38- haematopoietic cells, which when overexpressed, enforces quiescence in this population (Ellisen et al., 2001). *WT1* expression is decreased following RUNX1-EVI-1 knockdown. In AML, WT1 has both oncogenic and tumour suppressor characteristics in that, loss of function mutations are commonly seen, but simultaneously, overexpression of *WT1* is associated with a poor prognosis (Rampal and Figueroa, 2016). *WT1* encodes a zinc finger transcription factor which has been reported to direct TET2 to target genes. Mutations in WT1 disrupt this association with TET2 and leads to hypermethylation of target genes (Wang et al., 2015, Rampal et al., 2014).

In summary, the close association of t(3;21) leukaemia with early stem and progenitor cells is reflected in the expression of many genes important in programs important for the maintenance of the normal HSC compartment but is also frequently co-opted by leukaemic stem cells in order to share many functions associated with stem cells, such as self- renewal. A smaller subset of these genes are also activated in t(8;21) leukaemia, and are likely also to be part of the means by which this leukaemia maintains its self-renewal properties. The difference in expression of these stem cell genes are likely to contribute to the differences in prognoses between these two forms of leukaemia.

4.2 RUNX1-EVI-1 and RUNX1-ETO bind DNA in association with different transcription factor complexes

In this study we have performed a genome wide binding study of RUNX1-EVI-1, which is constitutively expressed in a t(3;21) cell line and have compared this binding pattern to RUNX1-ETO in a t(8;21) Kasumi-1 cell line. RUNX1-ETO DNA binding sites has been demonstrated previously in vivo, in both primary patient material and cell lines with the t(8;21) translocation (Ptasinska et al., 2012, Martens et al., 2012). However, DNA binding activity for RUNX1-EVI-1 has only been previously demonstrated in in vitro studies and transfected cell lines in which the fusion protein has been over expressed, therefore, this is the

first study to be able to draw a comparison between binding sites of both CBF fusion proteins.

In figure 3-5 and 3-6 we show that RUNX1-ETO and RUNX1-EVI-1 bind a substantial number of sites unique to each fusion protein. This difference in the repertoire of gene targets has important implications in terms of the mechanisms by which the two fusion proteins lead to leukaemia. One possible reason for the difference in binding sites may be differences in DNA binding motifs directly bound by each CBF fusion protein. RUNX1-ETO has only one DNA binding domain and thus can directly bind RUNX motifs. The three potential DNA binding domains (RUNT and two zinc-finger domains from the EVI-1 fusion, figure 1-3) may have additional DNA specificity. It has been previously shown only in in vitro studies that EVI-1 binds to either a GA(C/T)AAGA(T/C)AAGATAA (Delwel et al., 1993) or TGACAAGATAA sequence (Perkins et al., 1991). We were unable to identify either of these sequences in our motif searches, even we extended the sequence length criteria to accommodate this long motif. Although GATA motif containing sites are found in sites specifically bound by RUNX1-EVI-1 as compared to RUNX1-ETO, the binding of the EVI-1 partner to GATA motif alone would not explain why sites containing these motifs are also bound by RUNX1 in the t(3;21) leukaemia (figure 3-6 and 3-18). Further, in support of our argument that the EVI-1 part of RUNX1-EVI-1 does not bind DNA directly, others have shown that EVI-1 does not directly bind to DNA in a different cell line, but requires an interaction with GATA1 to do so. Mutations in EVI-1 that abolish this interaction prevent the functional effects of EVI-1 in retroviral overexpression in murine bone marrow cells (Laricchia-Robbio et al., 2006).

The differences in DNA binding sites are likely to be accounted for by the formation of different transcription factor complexes by each CBF fusion protein as a result of differing transcription factor interactions with either RUNX1-ETO or RUNX1-EVI-1 (figure 3-7C-D). We used a DNase-seq footprinting algorithm to enable the identification of motifs directly occupied by transcription factors in patient AML samples. By analysing DNA sequences bound by each CBF fusion

protein, as identified in ChIP-seq experiments in each cell line, in patient AML samples, we were able to identify which transcription factor families were associated with RUNX1-EVI-1 or RUNX1-ETO in purified primary, CD34+ AML cells. The occupied motifs identified in the patient samples were consistent with the enrichment of motifs in specific sites bound by each CBF fusion protein in the cell line (figure 3-6). We sampled this footprinted motif with a bootstrapping analysis in order to identify significantly co-localising transcription factors. This identified co-localising motifs occupied by PU.1-ERG-AP1-GATA transcription factors in RUNX1-EVI-1 binding sites, and RUNX-ERG-E-Box occupied motifs in RUNX1-ETO binding sites.

ETS transcription factor family member binding cooperativity is likely to be shared by the two CBF fusion proteins as this is promoted through the RUNX1 part of the fusion protein (Wotton et al., 1994). This may explain the presence of ERG occupied sites in both RUNX1-EVI-1 and RUNX1-ETO binding sites. The cooperativity between ERG and RUNX1-ETO binding has been previously shown (Martens et al., 2012).

In contrast, RUNX1-ETO but not RUNX1-EVI-1 forms a complex with E-Box binding transcription factors such as HEB and E2A (Sun et al., 2013, Ptasinska et al., 2014). This interaction is mediated by the NHR domain which is part of ETO, and the E protein (Sun et al., 2013), and hence this motif is specifically found at sites bound by RUNX1-ETO.

EVI-1, the fusion partner of RUNX1-EVI-1 has been shown to directly interact with the AP-1 family member FOS in a number of different cell lines (Bard-Chapeau et al., 2012). This interaction may be reinforced by the ability of EVI-1 to directly stimulate AP-1 signalling (Tanaka et al., 1994). In t(3;21) leukaemia, the most likely GATA factor to bind with RUNX1-EVI-1 is GATA2. GATA1 and 2 are the only GATA transcription factors expressed in both t(8;21) and t(3;21) leukaemia, however *GATA2* is the most likely candidate in terms of expression patterns as *GATA1* is mainly expressed in erythroid and megakaryocytic cells (figure3-7E)(Akashi et al., 2000). Furthermore, GATA2 binding sites from a related cell line are shared with both RUNX1-EVI-1 and RUNX1 (figure 3-7F).

This is in keeping with the observation that GATA motifs are found in binding sites of both RUNX1 and RUNX1-EVI-1 (figure 3-6C). Indeed, the loss of *GATA2* expression (but not *GATA1*) is seen following RUNX1-EVI-1 knockdown (figure 3-9E), and this correlates with a reduction in accessibility of GATA motif containing DHSs (figure 3-14B) which are sites strongly bound by GATA2 in the related leukaemic cell line (figure 3-18 B). This result is consistent with the observation that HSCs are extremely sensitive to *GATA2* mRNA levels. *Gata2*-/- mice cannot survive due to a failure of haematopoiesis (Tsai et al., 1994). Mice with heterozygous loss of *Gata2* can survive but show defects in fetal (Ling et al., 2004) and adult haematopoiesis (Rodrigues et al., 2005). In contrast murine HSC which overexpress *GATA2* by 2-fold display a block in their differentiation (Persons et al., 1999). The interaction between GATA2 and RUNX1-EVI-1 may be mediated through its RUNX1 domain. RUNX1 binding sites co-localise with the footprinted GATA motifs in t(3;21) leukaemic cells (figure 3-7A) and RUNX1 have been shown to physically interact with GATA2 in previous studies (Wilson et al., 2010).

4.2.1 CBF fusion protein binding pattern is intrinsically linked to the DHS landscape

GATA motifs are frequently found in DHSs unique to t(3;21) leukaemia, whilst, E-Box motifs are frequently found in DHSs unique to t(8;21) leukaemia (figure 3-2). This directly reflects the predominance of these motifs found in analysis of the CBF fusion protein binding sites (figure 3-5 and 3-6). These unique t(3;21) DHSs are likely bound by a complex containing ERG/AP1/PU.1/RUNX1-EVI-1/GATA2, whilst the unique t(8;21) DHSs are likely bound by an ERG/RUNX1-ETO/(HEB/E2A) complex, as described above. For example, one of these GATA motif containing sites unique to t(3;21) as opposed to t(8;21) leukaemia lies at an intragenic location within *MEIS1* (figure 3-1A) and is bound by both RUNX1 and RUNX1-EVI-1 (Figure 3-4 F), as well as GATA2 in the TF1 cell line (data not shown).

One interpretation of this observation would be that the different DHS landscape in each leukaemia allows a permissive environment for the binding of the

different transcription factor complexes. However, evidence is also present that supports the interpretation that these transcription factor complexes are important in the maintenance of this DHS landscape specific to each leukaemia. Following RUNX1-EVI-1 knockdown, the reduction in *GATA2* expression levels are associated with decrease in accessibility of a subset of GATA motif containing DHSs. This subset of DHSs is of significance because it is associated with decreased expression of neighbouring genes (figure3-14 A-B). .

This interaction between each CBF fusion protein complex and the DHS landscape is reinforced by differential cis-regulation and expression of members of the complex, such as GATA2. *GATA2* is more highly expressed in t(3;21) cells as compared to t(8;21) cells. This is driven by differences in accessibility of cis-regulatory elements in this gene. A distal *GATA2* enhancer, known to regulate *GATA2* expression, is accessible in normal CD34+ cells and t(8;21) patient cells but not in t(3;21) cells (Groschel et al., 2014). The *GATA2* promoter, is accessible and bound by RUNX1 in t(3;21) but not t(8;21) cells. Interestingly, in murine HSC *Gata2* promoter is thought to be bound by EVI-1 (Yuasa et al., 2005), however, in our human leukaemic cells, RUNX1-EVI-1 does not appear to bind at the promoter, but in an intragenic site. Notably, *GATA2* expression falls following knockdown of RUNX1-EVI-1. ERG is a common member of both RUNX1-EVI-1 and RUNX1-ETO complexes and is directly regulated by both CBF fusion proteins, as demonstrated by ChIP-seq and that expression of *ERG* decrease following knockdown of either CBF fusion proteins.

In summary, both RUNX1 and the CBF fusion protein complexes bind cis-regulatory modules, unique to each leukaemia, that through auto-regulation of its members, allow a stable gene regulatory network to be formed, which ultimately defines the behaviour of each leukaemia (Pimanda and Gottgens, 2010, Pimanda et al., 2007).

4.3 Differing roles of wild-type RUNX1 in t(3;21) and t(8;21) leukaemia

The untranslocated RUNX1 plays a vital role in RUNX1-ETO driven AML and the biological implication and mechanism for this has been carefully elucidated. Knockdown of the untranslocated RUNX1 leads to apoptosis of t(8;21) cells (Ben-Ami et al., 2013). The untranslocated RUNX1 competes with RUNX1-ETO for transcription factors in order to form complexes for binding to overlapping genomic loci. Knockdown of RUNX1-ETO leads to increased binding of RUNX1 at sites previously bound by RUNX1-ETO, replacing binding of repressive complexes with gene activating co-factors (Ptasinska et al., 2014).

The role of RUNX1 in the context of RUNX1-EVI-1 has not been fully examined before, but there is some evidence that there is also a dominant negative effect of RUNX1-EVI-1 towards native RUNX1, both on a direct protein-protein interaction and DNA binding capacity (Introduction 1.7.3). In our study we demonstrate there is a 74% overlap in RUNX1 and RUNX1-EVI-1 binding sites (figure 3-6A). It appears that the ability to also target genes through the RUNX1 part is crucial to the function of the RUNX1-EVI-1 fusion protein. When EVI-1 has been overexpressed in RUNX1 -/- murine bone marrow cells (Takeshita et al., 2008) transforming properties were not detectable by colony replating assays. The experiment where the full length RUNX1-EVI-1 molecule was transfected into RUNX1-/- cells was not performed to confirm if wild type RUNX1 was required for leukemogenecity of RUNX1-EVI-1. However, we do not necessarily demonstrate an active competition between RUNX1-EVI-1 and native RUNX1 for DNA binding sites as following RUNX1-EVI-1 knockdown we do not observe an upregulation of RUNX1 binding at sites previously bound by RUNX1-EVI-1 (figure 3-17B). The reason for this could be that at RUNX1-EVI-1 bound sites, RUNX binding motifs are not necessarily occupied, and that binding at these sites by RUNX1-EVI-1 is predominantly facilitated through the binding of other transcription factor partners. In keeping with this the RUNX footprinted motif is conspicuously absent in the core-complex at RUNX1EVI-1

bound sites (figure 3-7C) but is present in the complex at sites bound by RUNX1 (figure 3-7A).

4.4 RUNX1-EVI-1 knockdown and overexpression provides functional and mechanistic insight into the t(3;21) leukaemia

4.4.1 Maintenance of RUNX1-ETO, but not RUNX1-EVI-1 expressing CD34+ PBSCs

Although we were able to sustain RUNX1-ETO transduced CD34+ PBSCs, we were not able to do this for CD34+ PBSCs transduced with RUNX1-EVI-1 (figure3-8C). Although this could have been in part due to the poorer MOI achieved with this virus due to the size of the transgene, this was unlikely to be the sole cause, as even with equivalent MOI of virus, we obtained lower levels of infected CD34+ cells and cells which were transduced died early. RUNX1-EVI-1 is often seen as a secondary mutation, in contrast to RUNX1-ETO, either in CML, MDS which progresses to AML or in secondary AML following previous chemotherapy (Rubin et al., 1990, Rubin et al., 1987) (section 1.7.3). Therefore, RUNX1-EVI-1 may require co-operation with other mutations in order for the transformed cells to survive. Indeed, surprisingly, knockdown of RUNX1-EVI-1 leads to an early upregulation of *BCL2*, which is a direct target of RUNX1-EVI-1 (figure 3-12) and this is associated with an increase, although not statistically significant, in growth at the day 2 post RUNX1-EVI-1 knockdown (figure 3-10 C).

Although RUNX1-EVI-1 may cooperate with other mutations, it should still be possible to over-express RUNX1-EVI-1. This is because RUNX1-EVI-1 can be sustainably expressed independently in other transgenic animal models where either all cells express RUNX1-EVI-1 in an inducible manner (zebrafish) (Shen et al., 2015) or is expressed under the RUNX1 promoter (mouse) (Maki et al., 2005) or through retroviral overexpression in murine bone marrow cells (Cuenco et al., PNAS, 2000). It may be that RUNX1-EVI-1 transformation is dependent on the phenotype of the target cell. For example, RUNX1-EVI-1 transformation may require a more immature, less differentiated cell type.

Although we purified our apheresis material to select CD34+ PBSCs, this is likely to contain a heterogeneous population of LT-HSCs, ST-HSCs, MPPs, GMP etc. There are indeed many examples in the literature that suggests transforming cells at different stages of differentiation with oncogenes lead to different outcomes; these include the MLL fusion proteins (Krivtsov et al., 2013, Stavropoulou et al., 2016) and mutant members of the cohesin complex (Mazumdar et al., 2015).

Similarly, our results of transducing RUNX1-ETO into CD34+ PBSCs fail to maintain their CD34+ surface expression (figure 3-8G) and this is in keeping with another group's attempt to transduce CD34+ PBSCs with RUNX1-ETO (Wichmann et al., 2015). However, when others have transduced CD34+ cord blood cells with RUNX1-ETO, these cells maintain the expression of the immature surface marker CD34 (Mulloy et al., 2002), suggesting that the differences in cell population between even CD34+ PBSCs and cord blood is sufficient to produce a difference in phenotype following retroviral transduction of oncogenes.

4.4.2 RUNX1-EVI-1 maintains an aggressive leukaemic phenotype in t(3;21) cells

Our data showing that RUNX1 EVI-1 knockdown results in changes in differentiation of the t(3;21) cells is supported by the dysplastic phenotype of mice with RUNX1-EVI-1 knocked-in to the RUNX1 locus (Maki et al., 2005). In RUNX1-EVI-1 transgenic mice, definitive haematopoiesis fails to develop, and primitive haematopoietic cells in the fetal liver fail to differentiate into erythroid cells. In zebrafish, the expression of RUNX1-EVI-1 under a inducible heat shock promoter led to an accumulation of early myeloid progenitors and blast like cells (Shen et al., 2015). Similar to our results, RUNX1-EVI-1 expressing cells also showed increase proliferation and resistance to apoptosis.

We have already discussed a number of mechanisms by which this stem cell phenotype may be maintained. Other important RUNX1-EVI- target genes that are downregulated in SKH-1 following RUNX1-EVI-1 knockdown are *MSI2* and

ZEB1. *MSI2* (also known as Musashi) is a RNA binding protein and is a direct target of RUNX1-EVI-1 in our cells. *MSI2* expression is high in cells with CML in blast crisis and associated with a poor prognosis in these patients. MSI2 have been shown to inhibit Numb signalling pathway by binding to the 3'UTR of the transcript. Loss of *Msi2* expression or overexpression of *Numb* abrogates the development of blasts crisis in a murine model of CML (Ito et al., 2010).

ZEB1 is another gene that is directly targeted by RUNX1-EVI-1 and is downregulated following RUNX1-EVI-1 knockdown. The epithelial mesenchymal transition is the process during oncogenic transformation by which normal tissues lose their polarity and adhesion to neighbouring cells, thereby allowing metastasis to occur (Tam and Weinberg, 2013). ZEB1 is a transcription factor shown to play a critical role in the epithelial-mesenchymal transition of solid tumours (Chaffer et al., 2013). The role of ZEB1 has been recently shown in a MLL-AF9 model of leukaemia (Stavropoulou et al., 2016). *Zeb1* expression was critical for newly transformed leukaemic cells to depart from its bone marrow niche and infiltrate other organs. RUNX1-EVI-1 therefore targets and maintains a gene program that maintains the aggressive phenotype of this leukaemia.

4.5 The down regulation of C/EBPα by both CBF fusion proteins is critically required to maintain their leukaemic phenotype

Despite the differences between t(3;21) and t(8;21) leukaemia, C/EBPα is downregulated in both types of leukaemia (figure 3-16) suggesting that it is a critical node by which leukaemia is maintained. We show that C/EBPα may be directly repressed by both RUNX1-EVI-1 and RUNX1-ETO through binding of a recently characterised upstream enhancer (figure 3-16 D) (Avellino et al., 2016). Given the importance of inhibiting C/EBPα, there may be multiple means by which this is achieved (Pabst et al., 2001, Helbling et al., 2004). Others have suggested that RUNX1-EVI-1 may inhibit C/EBPα (introduction section 1.7.3) both through physical interaction as well as inhibition of a chaperone required to maintain correct mRNA levels of C/EBPα.

In our SKH-1 cell line, knockdown of RUNX1-EVI-1 leads to upregulation of C/EBPα and C/EBPδ (figure 3-15 C). However, C/EBPα is likely to be the more important factor for a number of reasons. Our RNA-seq data suggests that although C/EBPδ levels also increase, they are far lower than those of C/EBPα (figure 3-15 C). Our experiments involving the inhibition of CEPB transcription factors by a dominant negative binding partner are capable of inhibiting both C/EBPα and C/EBPδ. However, our ChIP-qPCR data directly show the binding of C/EBPα is affected by this DNCEPB peptide at key genes required for myeloid function (figure 3-21). Our C/EBPα overexpression data (figure 3-22 and 3-23) specifically show that C/EBPα overexpression phenocopies RUNX1-EVI-1 knockdown both at a gene expression level, as well as in phenotypic changes. C/EBPα but not C/EBPδ is consistently downregulated in primary AML patients with CBF translocations as compared to normal karyotype AML (figure 3-16). Finally, C/EBPδ knockout mice have little overt phenotype (Tanaka et al., 1997, Sterneck et al., 1998), in contrast C/EBPα knockout mice fail to develop mature neutrophils and eosinophils (Zhang et al., 1997a) with a block at the development of GMP from CMP (Zhang et al., 2004b).

4.5.1 C/EBPα binding leads to increase accessibility at large numbers of DHSs at genes required for myeloid differentiation

Given the inability of C/EBPα knockout mice to develop neutrophils it is perhaps not surprising that in our system blocking C/EBPα DNA binding by DNCEBP, following RUNX1-EVI-1 knockdown, entirely abolishes the upregulation of genes required for terminal myeloid function (*MPO, CSF1R* and *CTSG*) (figure 3-20). These genes are also upregulated when we overexpress C/EBPα in these cells (figure 3-23). We show that following knockdown of RUNX1-EVI-1, 2510 DHS increase in accessibility, and these are sites where the CEBP motif is frequently found. This result is associated with the increased binding of C/EBPα and RUNX1 at these sites (figure 3-18) following RUNX1-EVI-1knockdown. We show that at a number of these genes (*MPO, CTSG, TREM1* and *SIGLEC1)*, many of which are required for neutrophils and monocytes, the blocking of C/EBPα binding prevents the recruitment of RUNX1 and prevents

changes in the chromatin accessibility at sites which are likely to be cis-regulatory regions for these genes.

The ability of C/EBPα to alter chromatin accessibility is consistent with other studies that show that C/EBPα can interact with SWI/SNF nucleosome remodelling complexes (introduction 1.1.3) and that this interaction is important for the development of adipocytes (Pedersen et al., 2001). This ability to initiate the re-programming of chromatin structures may be the reason why C/EBPα is able to trans-differentiate cells committed to other lineages to myeloid cells (introduction section 1.1.3) and C/EBPα has also been shown to be important in the initiation of MLL-ENL leukaemia, likely to be through increasing accessibility of chromatin for the fusion protein to bind to target genes (Ohlsson et al., 2014).

4.5.2 C/EBPα binding is associated with the inhibition of stem cell gene expression program

We also observe that blocking C/EBPα binding prevents downregulation of CD34 following RUNX1-EVI-1 knockdown (figure 3-20 B) and in turn, overexpression of C/EBPα is sufficient to down-regulate CD34 despite the continued presence of RUNX1-EVI-1 (figure 3-22).

Overexpression of C/EBPα is also sufficient to downregulate stem cell associated genes such as *HOXA9, GATA2* and *MEIS1* (figure 3-23). The effects of DNCEBP on changes in stem cell associated gene expression in SKH-1 cells following RUNX1-EVI-1 knockdown appears to be more selective: although *HOXA9* downregulation is completely abolished, the effects on *MEIS1* and *GATA2* are relatively mild (figure 3-20D). The difference between the C/EBPα overexpression and DNCEBP experiments maybe that the level of C/EBPα overexpression may displace other transcription factor binding in the cells. For example, following RUNX1-EVI-1 knockdown we observe an increase in C/EBPα binding at sites previously bound by RUNX1-EVI-1 (figure 3-17B), therefore, it maybe that with sufficient levels of C/EBPα, this may be able to displace other transcription factors at other sites. In the DNCEBP system, the downregulation of RUNX1-EVI-1 may be sufficient to cause the

downregulation of some key stem cell specific genes, independently of the effects of C/EBPα. In keeping with this, *MEIS1* and *GATA2* (figure3-4F) but not *HOXA9* (data not shown) appears to be direct targets of RUNX1-EVI-1.

Consistent with this idea HSCs deficient in C/EBPα have increased competitive repopulating capacity and self-renewal (Zhang et al., 2004b). Although our demonstration that C/EBPα binding is required to "switch off" *HOXA9* appears to be novel, this is similar to the requirement for C/EBPα to switch off *mycn* in murine HSC development (Ye et al., 2013). In absence of C/EBPα, murine adult HSCs resemble fetal HSCs, in terms of a high proliferative capacity. C/EBPα binds *mycn* and represses this gene in adult HSCs which leads to quiescence. In our system, *MYCN* is also downregulated following RUNX1-EVI-1 knockdown, and is bound by C/EBPα (figure 3-17D). Other examples of genes associated with self-renewal capacity that are directly targeted by C/EBPα and repressed following RUNX1-EVI-1 knockdown include *SOX4* and *BMI-1* (Zhang et al., 2004b, Zhang et al., 2013).

4.6 Summary

In this work, we have demonstrated that RUNX1-EVI-1 and RUNX1-ETO interact with a different epigenetic landscape in each type of AML through different transcription factor partners. However, through knockdown of both RUNX1-EVI-1, and RUNX1-ETO (based on data previously published by our lab) we identify a core set of genes maintained by both fusion proteins. This core set of genes are likely to represent genes which are commonly deregulated amongst many leukaemia types. One of these genes is *CEBPA* which we demonstrate is commonly downregulated amongst all CBF leukaemias and is a critical node by which the transcription factor network is deregulated.

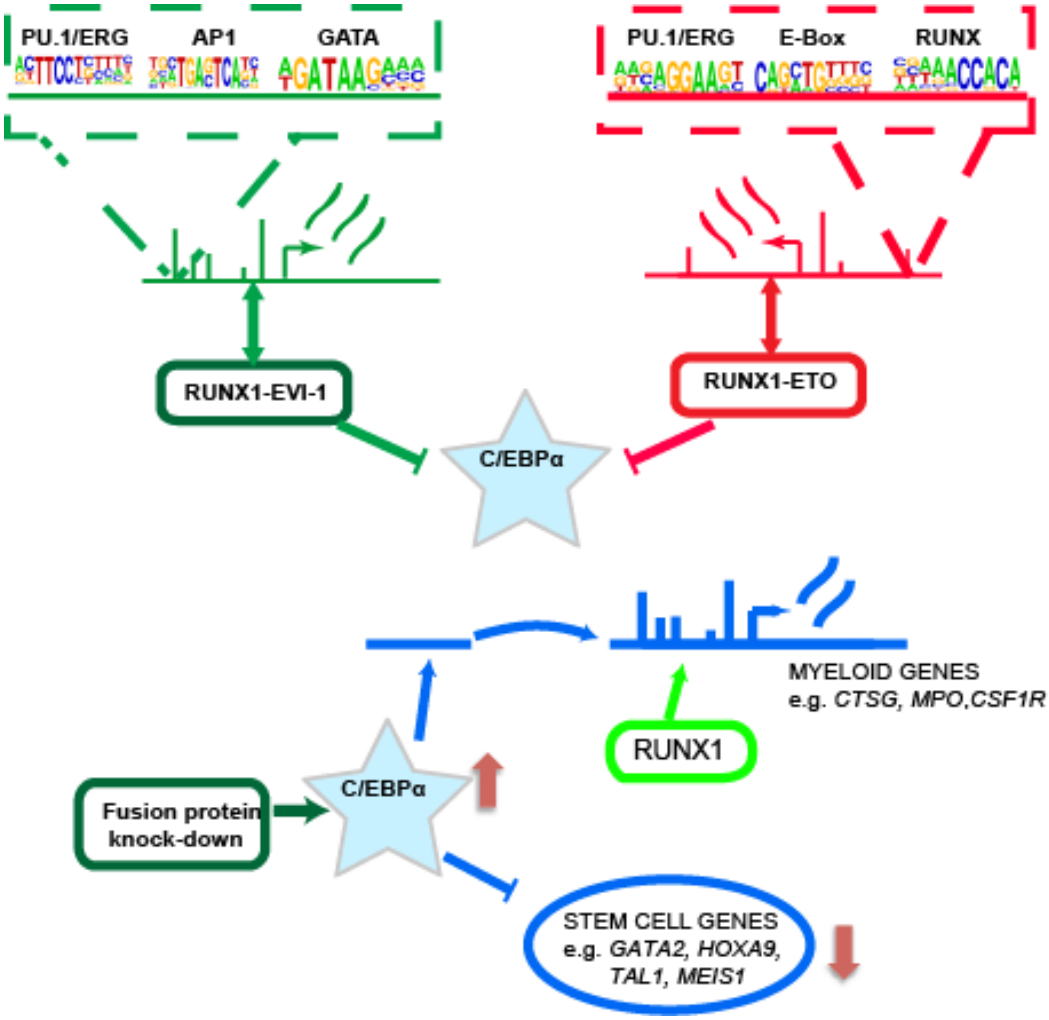


Figure 4-1 Graphical summary
RUNX1-EVI-1 and RUNX1-ETO differentially program the epigenetic landscape through different DNA binding partners. However, depletion of either CBF fusion proteins reliefs the inhibition of C/EBPα expression, resulting in direct inhibition of stem cell required genes and upregulation of genes required for myelopoiesis.

4.7 Future work

4.7.1 Ectopic expression of RUNX1-EVI-1 at varying stages of haematopoiesis

We have shown that RUNX1-EVI-1 leukaemia appear to be associated with the expression of many genes which are characteristic of both HSCs and LSCs. It is unclear to what extent this is enforced or simply maintained by RUNX1-EVI-1. Ectopic expression of this fusion gene in either an inducible transgenic model

(Stavropoulou et al., 2016) at different stages of haematopoietic development or through retroviral induction of different populations isolated from the haematopoietic hierarchy (Mazumdar et al., 2015), would decisively address this question.

4.7.2 What are the shared epigenetic mechanisms between the two CBF fusion proteins?

Both C-terminal fusion partners can recruit repressive complexes, albeit differing in identity. Both CBF fusion proteins can recruit histone deacetylases: RUNX1-ETO through NCoR/Sin3A and RUNX1-EVI-1 through CtBP. However, a distinguishing property of RUNX1-EVI-1 may be the formation of H3K9me3: it appears that CtBP and EVI-1 in its native form can recruit H3K9 methyltransferases such as G9a (Chinnadurai, 2007, Goyama et al., 2010), and EVI-1 itself can methylate H3K9 to form heterochromatin (Pinheiro et al., 2012). This in turn can recruit DNA methyl transferases such as DNMT1 (Esteve et al., 2006): DNA methylation is known to interact with H3K9 methyltransferases to re-inforce a silenced chromatin structure. Future work could involve identification of these partner proteins that may explain how certain RUNX1-EVI-1 target genes increase in expression, whilst others decrease in expression, following RUNX1-EVI-1 knockdown (figure3-12 A).

4.7.3 Role of differing CEBP transcription factors following RUNX1-EVI-1 knockdown

Despite the evidence that C/EBPα is the key CEBP transcription factor that is repressed by RUNX1-EVI-1, C/EBPδ levels also increase following RUNX1-EVI-1 knockdown, and our DNCEBP peptide is capable of inhibiting the DNA binding potential of both transcription factors. At present it is unclear as to the extent of the role of C/EBPδ following RUNX1-EVI-1 knockdown, begging the question of what would be the consequences of overexpressing this transcription factor. Furthermore, it is unclear to what extent would C/EBPδ compensate in our system if C/EBPα was knocked down in combination with RUNX1-EVI-1 knockdown.

4.7.4 Therapeutic targeting of C/EBPα expression in CBF patients

Previous groups have shown that C/EBPα trans-differentiation of lymphocytes lead to an irreversible switch to the myeloid lineage, even when the inducible C/EBPα is subsequently withdrawn (Bussmann et al., 2009). This is in keeping with evidence that demonstrates that once C/EBPα mediates the transition from CMP to GMP it is not required thereafter (Zhang et al., 2004b). It would be interesting to note how far this parallels the expression of C/EBPα in SKH-1 cells. In our system, although these CEBPA-ER SKH-1 cells upregulate surface expression of CD11b (figure 3-23D-E), they rapidly undergo apoptosis. It is unclear whether these cells would otherwise be sustainable if cultured with cytokines, such as G-CSF. Indeed given the vulnerability of both t(3;21) cells and t(8;21) cells (Pabst et al., 2001) to C/EBPα overexpression, this may provide an unifying therapeutic strategy for both forms of leukaemia. Overexpression of C/EBPα may be delivered therapeutically through a short-activating RNA sequence that act on the promoter of C/EBPα, and this has shown promise in the treatment of hepatocellular carcinoma (Reebye et al., 2014).

References

ADOLFSSON, J., MÅNSSON, R., BUZA-VIDAS, N., HULTQUIST, A., LIUBA, K., JENSEN, C. T., BRYDER, D., YANG, L., BORGE, O.-J., THOREN, L. A. M., ANDERSON, K., SITNICKA, E., SASAKI, Y., SIGVARDSSON, M. & JACOBSEN, S. E. W. 2005. Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. *Cell,* 121**,** 295-306.

AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature,* 404**,** 193-197.

ALLFREY, V. G., FAULKNER, R. & MIRSKY, A. E. 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc Natl Acad Sci U S A,* 51**,** 786-794.

ANDREEFF, M., RUVOLO, V., GADGIL, S., ZENG, C., COOMBES, K., CHEN, W., KORNBLAU, S., BARON, A. E. & DRABKIN, H. A. 2008. HOX expression patterns identify a common signature for favorable AML. *Leukemia,* 22**,** 2041-7.

ARIAS, J., ALBERTS, A. S., BRINDLE, P., CLARET, F. X., SMEAL, T., KARIN, M., FERAMISCO, J. & MONTMINY, M. 1994. Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature,* 370**,** 226-9.

ARINOBU, Y., MIZUNO, S.-I., CHONG, Y., SHIGEMATSU, H., IINO, T., IWASAKI, H., GRAF, T., MAYFIELD, R., CHAN, S., KASTNER, P. & AKASHI, K. 2007. Reciprocal Activation of GATA-1 and PU.1 Marks Initial Specification of Hematopoietic Stem Cells into Myeloerythroid and Myelolymphoid Lineages. *Cell Stem Cell,* 1**,** 416-427.

ASOU, H., TASHIRO, S., HAMAMOTO, K., OTSUJI, A., KITA, K. & KAMADA, N. 1991. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood,* 77**,** 2031-6.

AVELLINO, R., HAVERMANS, M., ERPELINCK, C., SANDERS, M. A., HOOGENBOEZEM, R., VAN DE WERKEN, H. J., ROMBOUTS, E., VAN LOM, K., VAN STRIEN, P. M., GEBHARD, C., REHLI, M., PIMANDA, J., BECK, D., ERKELAND, S., KUIKEN, T., DE LOOPER, H., GROSCHEL, S., TOUW, I., BINDELS, E. & DELWEL, R. 2016. An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation. *Blood*.

BALDRIDGE, M. T., KING, K. Y., BOLES, N. C., WEKSBERG, D. C. & GOODELL, M. A. 2010. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature,* 465**,** 793-7.

BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

BANNISTER, A. J., ZEGERMAN, P., PARTRIDGE, J. F., MISKA, E. A., THOMAS, J. O., ALLSHIRE, R. C. & KOUZARIDES, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature,* 410**,** 120-124.

BARCLAY, J., BALAGUERO, N., MIONE, M., ACKERMAN, S. L., LETTS, V. A., BRODBECK, J., CANTI, C., MEIR, A., PAGE, K. M., KUSUMI, K., PEREZ-REYES, E., LANDER, E. S., FRANKEL, W. N., GARDINER, R. M., DOLPHIN, A. C. & REES, M. 2001. Ducky mouse phenotype of epilepsy and ataxia is associated with mutations in the

Cacna2d2 gene and decreased calcium channel current in cerebellar Purkinje cells. *J Neurosci,* 21**,** 6095-104.

BARD-CHAPEAU, E. A., JEYAKANI, J., KOK, C. H., MULLER, J., CHUA, B. Q., GUNARATNE, J., BATAGOV, A., JENJAROENPUN, P., KUZNETSOV, V. A., WEI, C. L., D'ANDREA, R. J., BOURQUE, G., JENKINS, N. A. & COPELAND, N. G. 2012. Ecotopic viral integration site 1 (EVI1) regulates multiple cellular processes important for cancer and is a synergistic partner for FOS protein in invasive tumors. *Proc.Natl.Acad.Sci.U.S A,* 109**,** 2168-2173.

BARRAL, P., POLZELLA, P., BRUCKBAUER, A., VAN ROOIJEN, N., BESRA, G. S., CERUNDOLO, V. & BATISTA, F. D. 2010. CD169(+) macrophages present lipid antigens to mediate early activation of iNKT cells in lymph nodes. *Nat Immunol,* 11**,** 303-12.

BELOTSERKOVSKAYA, R., OH, S., BONDARENKO, V. A., ORPHANIDES, G., STUDITSKY, V. M. & REINBERG, D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science,* 301**,** 1090-1093.

BEN-AMI, O., FRIEDMAN, D., LESHKOWITZ, D., GOLDENBERG, D., ORLOVSKY, K., PENCOVICH, N., LOTEM, J., TANAY, A. & GRONER, Y. 2013. Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep,* 4**,** 1131-43.

BERNT, KATHRIN M., ZHU, N., SINHA, AMIT U., VEMPATI, S., FABER, J., KRIVTSOV, ANDREI V., FENG, Z., PUNT, N., DAIGLE, A., BULLINGER, L., POLLOCK, ROY M., RICHON, VICTORIA M., KUNG, ANDREW L. & ARMSTRONG, SCOTT A. 2011. MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. *Cancer cell,* 20**,** 66-78.

BINDEA, G., MLECNIK, B., HACKL, H., CHAROENTONG, P., TOSOLINI, M., KIRILOVSKY, A., FRIDMAN, W. H., PAGES, F., TRAJANOSKI, Z. & GALON, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics,* 25**,** 1091-3.

BINDELS, E. M., HAVERMANS, M., LUGTHART, S., ERPELINCK, C., WOCJTOWICZ, E., KRIVTSOV, A. V., ROMBOUTS, E., ARMSTRONG, S. A., TASKESEN, E., HAANSTRA, J. R., BEVERLOO, H. B., DOHNER, H., HUDSON, W. A., KERSEY, J. H., DELWEL, R. & KUMAR, A. R. 2012. EVI1 is critical for the pathogenesis of a subset of MLL-AF9 rearranged AMLs. *Blood*.

BOISSET, J.-C., VAN CAPPELLEN, W., ANDRIEU-SOLER, C., GALJART, N., DZIERZAK, E. & ROBIN, C. 2010. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature,* 464**,** 116-120.

BOMKEN, S., BUECHLER, L., REHE, K., PONTHAN, F., ELDER, A., BLAIR, H., BACON, C. M., VORMOOR, J. & HEIDENREICH, O. 2013. Lentiviral marking of patient-derived acute lymphoblastic leukaemic cells allows in vivo tracking of disease progression. *Leukemia,* 27**,** 718-721.

BONDARENKO, V. A., STEELE, L. M., ÚJVÁRI, A., GAYKALOVA, D. A., KULAEVA, O. I., POLIKANOV, Y. S., LUSE, D. S. & STUDITSKY, V. M. 2006. Nucleosomes Can Form a Polar Barrier to Transcript Elongation by RNA Polymerase II. *Molecular Cell,* 24**,** 469-479.

BONNET, D. & DICK, J. E. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med,* 3**,** 730-7.

BORREGAARD, N., SØRENSEN, O. E. & THEILGAARD-MÖNCH, K. 2007. Neutrophil granules: a library of innate immunity proteins. *Trends in Immunology,* 28**,** 340-345.

BOSTICK, M., KIM, J. K., ESTÈVE, P.-O., CLARK, A., PRADHAN, S. & JACOBSEN, S. E. 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science,* 317**,** 1760-1764.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell,* 132**,** 311-322.

BROWN, R. C., PATTISON, S., VAN REE, J., COGHILL, E., PERKINS, A., JANE, S. M. & CUNNINGHAM, J. M. 2002. Distinct Domains of Erythroid Krüppel-Like Factor Modulate Chromatin Remodeling and Transactivation at the Endogenous β-Globin Gene Promoter. *Molecular and Cellular Biology,* 22**,** 161-170.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Meth,* 10**,** 1213-1218.

BULTMAN, S. J., GEBUHR, T. C. & MAGNUSON, T. 2005. A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in β-globin expression and erythroid development. *Genes & Development,* 19**,** 2849-2861.

BUONAMICI, S., LI, D., CHI, Y., ZHAO, R., WANG, X., BRACE, L., NI, H., SAUNTHARARAJAH, Y. & NUCIFORA, G. 2004. EVI1 induces myelodysplastic syndrome in mice. *The Journal of Clinical Investigation,* 114**,** 713-719.

BUREL, S. A., HARAKAWA, N., ZHOU, L., PABST, T., TENEN, D. G. & ZHANG, D. E. 2001. Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation. *Mol Cell Biol,* 21**,** 5577-90.

BURNETT, A. K., RUSSELL, N. H., HILLS, R. K., KELL, J., CAVENAGH, J., KJELDSEN, L., MCMULLIN, M.-F., CAHALIN, P., DENNIS, M., FRIIS, L., THOMAS, I. F., MILLIGAN, D. & CLARK, R. E. 2015. A randomized comparison of daunorubicin 90 mg/m2 vs 60 mg/m2 in AML induction: results from the UK NCRI AML17 trial in 1206 patients. *Blood,* 125**,** 3878-3885.

BURNETT, A. K., RUSSELL, N. H., HILLS, R. K., KELL, J., FREEMAN, S., KJELDSEN, L., HUNTER, A. E., YIN, J., CRADDOCK, C. F., DUFVA, I. H., WHEATLEY, K. & MILLIGAN, D. 2012. Addition of gemtuzumab ozogamicin to induction chemotherapy improves survival in older patients with acute myeloid leukemia. *J Clin Oncol,* 30**,** 3924-31.

BUSCH, K., KLAPPROTH, K., BARILE, M., FLOSSDORF, M., HOLLAND-LETZ, T., SCHLENNER, S. M., RETH, M., HOFER, T. & RODEWALD, H. R. 2015. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature,* 518**,** 542-6.

BUSSMANN, L. H., SCHUBERT, A., VU MANH, T. P., DE ANDRES, L., DESBORDES, S. C., PARRA, M., ZIMMERMANN, T., RAPINO, F., RODRIGUEZ-UBREVA, J., BALLESTAR, E. & GRAF, T. 2009. A Robust and Highly Efficient Immune Cell Reprogramming System. *Cell Stem Cell,* 5**,** 554-566.

BYRD, J. C., MRÓZEK, K., DODGE, R. K., CARROLL, A. J., EDWARDS, C. G., ARTHUR, D. C., PETTENATI, M. J., PATIL, S. R., RAO, K. W., WATSON, M. S., KODURU, P. R. K., MOORE, J. O., STONE, R. M., MAYER, R. J., FELDMAN, E. J., DAVEY, F. R., SCHIFFER, C. A., LARSON, R. A. & BLOOMFIELD, C. D. 2002. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Presented in part at the 43rd annual meeting of the American Society of Hematology, Orlando, FL, December 10, 200,* 100**,** 4325-4336.

CABEZAS-WALLSCHEID, N. 2013. Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. *EMBO Mol Med,* 5**,** 1804-1820.

CANCER GENOME ATLAS RESEARCH, N. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med,* 368**,** 2059-74.

CARBONI, G. L., GAO, B., NISHIZAKI, M., XU, K., MINNA, J. D., ROTH, J. A. & JI, L. 2003. CACNA2D2-mediated apoptosis in NSCLC cells is associated with alterations of the intracellular calcium signaling and disruption of mitochondria membrane integrity. *Oncogene,* 22**,** 615-26.

CAREY, M., LI, B. & WORKMAN, J. L. 2006. RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. *Molecular Cell,* 24**,** 481-487.

CATTANEO, F. & NUCIFORA, G. 2008. EVI1 recruits the histone methyltransferase SUV39H1 for transcription repression. *Journal of Cellular Biochemistry,* 105**,** 344-352.

CAUCHY, P., JAMES, SALLY R., ZACARIAS-CABEZA, J., PTASINSKA, A., IMPERATO, MARIA R., ASSI, SALAM A., PIPER, J., CANESTRARO, M., HOOGENKAMP, M., RAGHAVAN, M., LOKE, J., AKIKI, S., CLOKIE, SAMUEL J., RICHARDS, STEPHEN J., WESTHEAD, DAVID R., GRIFFITHS, MICHAEL J., OTT, S., BONIFER, C. & COCKERILL, PETER N. 2015. Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. *Cell Reports,* 12**,** 821-836.

CHAFFER, C. L., MARJANOVIC, N. D., LEE, T., BELL, G., KLEER, C. G., REINHARDT, F., D'ALESSIO, A. C., YOUNG, R. A. & WEINBERG, R. A. 2013. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell,* 154**,** 61-74.

CHAKRABORTY, S., SENYUK, V., SITAILO, S., CHI, Y. & NUCIFORA, G. 2001. Interaction of EVI1 with cAMP-responsive element-binding protein-binding protein (CBP) and p300/CBP-associated factor (P/CAF) results in reversible acetylation of EVI1 and in co-localization in nuclear speckles. *J.Biol.Chem.,* 276**,** 44936-44943.

CHALLEN, G. A., SUN, D., JEONG, M., LUO, M., JELINEK, J., BERG, J. S., BOCK, C., VASANTHAKUMAR, A., GU, H., XI, Y., LIANG, S., LU, Y., DARLINGTON, G. J.,

MEISSNER, A., ISSA, J.-P. J., GODLEY, L. A., LI, W. & GOODELL, M. A. 2012. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet,* 44**,** 23-31.

CHEN, T., HEVI, S., GAY, F., TSUJIMOTO, N., HE, T., ZHANG, B., UEDA, Y. & LI, E. 2007. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat Genet,* 39**,** 391-396.

CHI, T. H., WAN, M., ZHAO, K., TANIUCHI, I., CHEN, L., LITTMAN, D. R. & CRABTREE, G. R. 2002. Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. *Nature,* 418**,** 195-9.

CHIAPPINELLI, K. B., STRISSEL, P. L., DESRICHARD, A., LI, H., HENKE, C., AKMAN, B., HEIN, A., ROTE, N. S., COPE, L. M., SNYDER, A., MAKAROV, V., BUDHU, S., SLAMON, D. J., WOLCHOK, J. D., PARDOLL, D. M., BECKMANN, M. W., ZAHNOW, C. A., MERGHOUB, T., CHAN, T. A., BAYLIN, S. B. & STRICK, R. 2015. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell,* 162**,** 974-86.

CHINNADURAI, G. 2007. Transcriptional regulation by C-terminal binding proteins. *Int J Biochem Cell Biol,* 39**,** 1593-607.

CHRISTENSEN, J. L., WRIGHT, D. E., WAGERS, A. J. & WEISSMAN, I. L. 2004. Circulation and chemotaxis of fetal hematopoietic stem cells. *PLoS Biol,* 2**,** E75.

CIRILLO, L. A., LIN, F. R., CUESTA, I., FRIEDMAN, D., JARNIK, M. & ZARET, K. S. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell,* 9**,** 279-89.

COCKERILL, P. N. 2011. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J,* 278**,** 2182-210.

COCKERILL, P. N., BERT, A. G., ROBERTS, D. & VADAS, M. A. 1999. The human granulocyte-macrophage colony- stimulating factor gene is autonomously regulated in vivo by an inducible tissue-specific enhancer. *Proceedings of the National Academy of Sciences,* 96**,** 15097-15102.

COLLINS, C. T. & HESS, J. L. 2015. Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets. *Oncogene,* 35**,** 1090-8.

CUENCO, G. M., NUCIFORA, G. & REN, R. 2000. Human AML1/MDS1/EVI1 fusion protein induces an acute myelogenous leukemia (AML) in mice: a model for human AML. *Proc.Natl.Acad.Sci.U.S A,* 97**,** 1760-1765.

CUENCO, G. M. & REN, R. 2001. Cooperation of BCR-ABL and AML1/MDS1/EVI1 in blocking myeloid differentiation and rapid induction of an acute myelogenous leukemia. *Oncogene,* 20**,** 8236-8248.

CUENCO, G. M. & REN, R. 2004. Both AML1 and EVI1 oncogenic components are required for the cooperation of AML1/MDS1/EVI1 with BCR/ABL in the induction of acute myelogenous leukemia in mice. *Oncogene,* 23**,** 569-579.

CUMANO, A., DIETERLEN-LIEVRE, F. & GODIN, I. 1996. Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura. *Cell,* 86**,** 907-16.

CUMANO, A., FERRAZ, J. C., KLAINE, M., DI SANTO, J. P. & GODIN, I. 2001. Intraembryonic, but not yolk sac hematopoietic precursors, isolated before

circulation, provide long-term multilineage reconstitution. *Immunity,* 15**,** 477-85.

DAGA, A., TIGHE, J. E. & CALABI, F. 1992. Leukaemia/Drosophila homology. *Nature,* 356**,** 484.

DAHL, R., IYER, S. R., OWENS, K. S., CUYLEAR, D. D. & SIMON, M. C. 2006. The transcriptional repressor Gfi-1 antagonizes PU.1 activity through protein-protein interaction. *Journal of Biological Chemistry*.

DAHL, R., WALSH, J. C., LANCKI, D., LASLO, P., IYER, S. R., SINGH, H. & SIMON, M. C. 2003. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBP[alpha] ratio and granulocyte colony-stimulating factor. *Nat Immunol,* 4**,** 1029-1036.

DAI, P., AKIMARU, H., TANAKA, Y., HOU, D. X., YASUKAWA, T., KANEI-ISHII, C., TAKAHASHI, T. & ISHII, S. 1996. CBP as a transcriptional coactivator of c-Myb. *Genes Dev,* 10**,** 528-40.

DAI, X. M., RYAN, G. R., HAPEL, A. J., DOMINGUEZ, M. G., RUSSELL, R. G., KAPP, S., SYLVESTRE, V. & STANLEY, E. R. 2002. Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood,* 99**,** 111-20.

DAIGLE, S. R., OLHAVA, E. J., THERKELSEN, C. A., BASAVAPATHRUNI, A., JIN, L., BORIACK-SJODIN, P. A., ALLAIN, C. J., KLAUS, C. R., RAIMONDI, A., SCOTT, M. P., WATERS, N. J., CHESWORTH, R., MOYER, M. P., COPELAND, R. A., RICHON, V. M. & POLLOCK, R. M. 2013. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood,* 122**,** 1017-1025.

DAWSON, M. A., PRINJHA, R. K., DITTMANN, A., GIOTOPOULOS, G., BANTSCHEFF, M., CHAN, W. I., ROBSON, S. C., CHUNG, C. W., HOPF, C., SAVITSKI, M. M., HUTHMACHER, C., GUDGIN, E., LUGO, D., BEINKE, S., CHAPMAN, T. D., ROBERTS, E. J., SODEN, P. E., AUGER, K. R., MIRGUET, O., DOEHNER, K., DELWEL, R., BURNETT, A. K., JEFFREY, P., DREWES, G., LEE, K., HUNTLY, B. J. & KOUZARIDES, T. 2011. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature,* 478**,** 529-33.

DEKOTER, R. P. & SINGH, H. 2000. Regulation of B Lymphocyte and Macrophage Development by Graded Expression of PU.1. *Science,* 288**,** 1439-1441.

DELWEL, R., FUNABIKI, T., KREIDER, B. L., MORISHITA, K. & IHLE, J. N. 1993. Four of the seven zinc fingers of the Evi-1 myeloid-transforming gene are required for sequence-specific binding to GA(C/T)AAGA(T/C)AAGATAA. *Mol Cell Biol,* 13**,** 4291-300.

DENNIS, M., RUSSELL, N., HILLS, R. K., HEMMAWAY, C., PANOSKALTSIS, N., MCMULLIN, M.-F., KJELDSEN, L., DIGNUM, H., THOMAS, I. F., CLARK, R. E., MILLIGAN, D. & BURNETT, A. K. 2015. Vosaroxin and vosaroxin plus low-dose Ara-C (LDAC) vs low-dose Ara-C alone in older patients with acute myeloid leukemia. *Blood,* 125**,** 2923-2932.

DING, L., LEY, T. J., LARSON, D. E., MILLER, C. A., KOBOLDT, D. C., WELCH, J. S., RITCHEY, J. K., YOUNG, M. A., LAMPRECHT, T., MCLELLAN, M. D., MCMICHAEL, J. F., WALLIS, J. W., LU, C., SHEN, D., HARRIS, C. C., DOOLING, D. J., FULTON, R. S.,

FULTON, L. L., CHEN, K., SCHMIDT, H., KALICKI-VEIZER, J., MAGRINI, V. J., COOK, L., MCGRATH, S. D., VICKERY, T. L., WENDL, M. C., HEATH, S., WATSON, M. A., LINK, D. C., TOMASSON, M. H., SHANNON, W. D., PAYTON, J. E., KULKARNI, S., WESTERVELT, P., WALTER, M. J., GRAUBERT, T. A., MARDIS, E. R., WILSON, R. K. & DIPERSIO, J. F. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature,* 481**,** 506-510.

DOGAN, N., WU, W., MORRISSEY, C. S., CHEN, K. B., STONESTROM, A., LONG, M., KELLER, C. A., CHENG, Y., JAIN, D., VISEL, A., PENNACCHIO, L. A., WEISS, M. J., BLOBEL, G. A. & HARDISON, R. C. 2015. Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin,* 8**,** 16.

DUNNE, J., CULLMANN, C., RITTER, M., SORIA, N. M., DRESCHER, B., DEBERNARDI, S., SKOULAKIS, S., HARTMANN, O., KRAUSE, M., KRAUTER, J., NEUBAUER, A., YOUNG, B. D. & HEIDENREICH, O. 2006. siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. *Oncogene,* 25**,** 6067-78.

DUPLOYEZ, N., MARCEAU-RENAUT, A., BOISSEL, N., PETIT, A., BUCCI, M., GEFFROY, S., LAPILLONNE, H., RENNEVILLE, A., RAGU, C., FIGEAC, M., CELLI-LEBRAS, K., LACOMBE, C., MICOL, J.-B., ABDEL-WAHAB, O., CORNILLET, P., IFRAH, N., DOMBRET, H., LEVERGER, G., JOURDAN, E. & PREUDHOMME, C. 2016. Comprehensive mutational profiling of core binding factor acute myeloid leukemia. *Blood,* 127**,** 2451-2459.

DYNAN, W. S. & TJIAN, R. 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell,* 35**,** 79-87.

ELLISEN, L. W., CARLESSO, N., CHENG, T., SCADDEN, D. T. & HABER, D. A. 2001. The Wilms tumor suppressor WT1 directs stage-specific quiescence and differentiation of human hematopoietic progenitor cells. *The EMBO Journal,* 20**,** 1897-1909.

EPPERT, K., TAKENAKA, K., LECHMAN, E. R., WALDRON, L., NILSSON, B., VAN GALEN, P., METZELER, K. H., POEPPL, A., LING, V., BEYENE, J., CANTY, A. J., DANSKA, J. S., BOHLANDER, S. K., BUSKE, C., MINDEN, M. D., GOLUB, T. R., JURISICA, I., EBERT, B. L. & DICK, J. E. 2011. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med,* 17**,** 1086-93.

ERICKSON, P., GAO, J., CHANG, K. S., LOOK, T., WHISENANT, E., RAIMONDI, S., LASHER, R., TRUJILLO, J., ROWLEY, J. & DRABKIN, H. 1992. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood,* 80**,** 1825-31.

ESSERS, M. A., OFFNER, S., BLANCO-BOSE, W. E., WAIBLER, Z., KALINKE, U., DUCHOSAL, M. A. & TRUMPP, A. 2009. IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature,* 458**,** 904-8.

ESTEVE, P. O., CHIN, H. G., SMALLWOOD, A., FEEHERY, G. R., GANGISETTY, O., KARPF, A. R., CAREY, M. F. & PRADHAN, S. 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev,* 20**,** 3089-103.

FALINI, B., MECUCCI, C., TIACCI, E., ALCALAY, M., ROSATI, R., PASQUALUCCI, L., LA STARZA, R., DIVERIO, D., COLOMBO, E., SANTUCCI, A., BIGERNA, B., PACINI, R., PUCCIARINI, A., LISO, A., VIGNETTI, M., FAZI, P., MEANI, N., PETTIROSSI, V., SAGLIO, G., MANDELLI, F., LO-COCO, F., PELICCI, P. G. & MARTELLI, M. F. 2005. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med,* 352**,** 254-66.

FATEMI, M., HERMANN, A., PRADHAN, S. & JELTSCH, A. 2001. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J Mol Biol,* 309**,** 1189-99.

FENAUX, P., MUFTI, G. J., HELLSTROM-LINDBERG, E., SANTINI, V., FINELLI, C., GIAGOUNIDIS, A., SCHOCH, R., GATTERMANN, N., SANZ, G., LIST, A., GORE, S. D., SEYMOUR, J. F., BENNETT, J. M., BYRD, J., BACKSTROM, J., ZIMMERMAN, L., MCKENZIE, D., BEACH, C. L. & SILVERMAN, L. R. 2009. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *The Lancet Oncology,* 10**,** 223-232.

FIGUEROA, M. E., ABDEL-WAHAB, O., LU, C., WARD, P. S., PATEL, J., SHIH, A., LI, Y., BHAGWAT, N., VASANTHAKUMAR, A., FERNANDEZ, H. F., TALLMAN, M. S., SUN, Z., WOLNIAK, K., PEETERS, J. K., LIU, W., CHOE, S. E., FANTIN, V. R., PAIETTA, E., LÖWENBERG, B., LICHT, J. D., GODLEY, L. A., DELWEL, R., VALK, P. J. M., THOMPSON, C. B., LEVINE, R. L. & MELNICK, A. 2010a. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer cell,* 18**,** 553-567.

FIGUEROA, M. E., LUGTHART, S., LI, Y., ERPELINCK-VERSCHUEREN, C., DENG, X., CHRISTOS, P. J., SCHIFANO, E., BOOTH, J., VAN PUTTEN, W., SKRABANEK, L., CAMPAGNE, F., MAZUMDAR, M., GREALLY, J. M., VALK, P. J. M., LÖWENBERG, B., DELWEL, R. & MELNICK, A. 2010b. DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia. *Cancer cell,* 17**,** 13-27.

FOLLOWS, G. A., TAGOH, H., RICHARDS, S. J., MELNIK, S., DICKINSON, H., DE WYNTER, E., LEFEVRE, P., MORGAN, G. J. & BONIFER, C. 2005. c-FMS chromatin structure and expression in normal and leukaemic myelopoiesis. *Oncogene,* 24**,** 3643-51.

FRANK, R., ZHANG, J., UCHIDA, H., MEYERS, S., HIEBERT, S. W. & NIMER, S. D. 1995. The AML1/ETO fusion protein blocks transactivation of the GM-CSF promoter by AML1B. *Oncogene,* 11**,** 2667-74.

FUNABIKI, T., KREIDER, B. L. & IHLE, J. N. 1994. The carboxyl domain of zinc fingers of the Evi-1 myeloid transforming gene binds a consensus sequence of GAAGATGAG. *Oncogene,* 9**,** 1575-81.

GAIDZIK, V. I., BULLINGER, L., SCHLENK, R. F., ZIMMERMANN, A. S., R+ÂCK, J., PASCHKA, P., CORBACIOGLU, A., KRAUTER, J., SCHLEGELBERGER, B., GANSER, A., SP+ÑTH, D., K++NDGEN, A., SCHMIDT-WOLF, I. G. H., G+ÂTZE, K., NACHBAUR, D., PFREUNDSCHUH, M., HORST, H. A., D+ÂHNER, H. & D+ÂHNER, K. 2011. RUNX1 Mutations in Acute Myeloid Leukemia: Results From a

Comprehensive Genetic and Clinical Analysis From the AML Study Group. *Journal of Clinical Oncology,* 29**,** 1364-1372.

GAMOU, T., KITAMURA, E., HOSODA, F., SHIMIZU, K., SHINOHARA, K., HAYASHI, Y., NAGASE, T., YOKOYAMA, Y. & OHKI, M. 1998. The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the MTG8(ETO) family. *Blood,* 91**,** 4028-37.

GEKAS, C., DIETERLEN-LIEVRE, F., ORKIN, S. H. & MIKKOLA, H. K. 2005. The placenta is a niche for hematopoietic stem cells. *Dev Cell,* 8**,** 365-75.

GELMETTI, V., ZHANG, J., FANELLI, M., MINUCCI, S., PELICCI, P. G. & LAZAR, M. A. 1998. Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol,* 18**,** 7185-91.

GENOVESE , G., KÄHLER , A. K., HANDSAKER , R. E., LINDBERG , J., ROSE , S. A., BAKHOUM , S. F., CHAMBERT , K., MICK , E., NEALE , B. M., FROMER , M., PURCELL , S. M., SVANTESSON , O., LANDÉN , M., HÖGLUND , M., LEHMANN , S., GABRIEL , S. B., MORAN , J. L., LANDER , E. S., SULLIVAN , P. F., SKLAR , P., GRÖNBERG , H., HULTMAN , C. M. & MCCARROLL , S. A. 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine,* 371**,** 2477-2487.

GENTLES, A. J., PLEVRITIS, S. K., MAJETI, R. & ALIZADEH, A. A. 2010. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA,* 304**,** 2706-15.

GERHARDT, T. M., SCHMAHL, G. E., FLOTHO, C., RATH, A. V. & NIEMEYER, C. M. 1997. Expression of the Evi-1 gene in haemopoietic cells of children with juvenile myelomonocytic leukaemia and normal donors. *Br.J.Haematol.,* 99**,** 882-887.

GILLILAND, D. G. & GRIFFIN, J. D. 2002. The roles of FLT3 in hematopoiesis and leukemia. *Blood,* 100**,** 1532-42.

GLASS, C., WUERTZER, C., CUI, X., BI, Y., DAVULURI, R., XIAO, Y.-Y., WILSON, M., OWENS, K., ZHANG, Y. & PERKINS, A. 2013. Global Identification of EVI1 Target Genes in Acute Myeloid Leukemia. *PLoS ONE,* 8**,** e67134.

GOARDON, N., MARCHI, E., ATZBERGER, A., QUEK, L., SCHUH, A., SONEJI, S., WOLL, P., MEAD, A., ALFORD, K. A., ROUT, R., CHAUDHURY, S., GILKES, A., KNAPPER, S., BELDJORD, K., BEGUM, S., ROSE, S., GEDDES, N., GRIFFITHS, M., STANDEN, G., STERNBERG, A., CAVENAGH, J., HUNTER, H., BOWEN, D., KILLICK, S., ROBINSON, L., PRICE, A., MACINTYRE, E., VIRGO, P., BURNETT, A., CRADDOCK, C., ENVER, T., JACOBSEN, S. E., PORCHER, C. & VYAS, P. 2011. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer cell,* 19**,** 138-52.

GOLUB, T. R., BARKER, G. F., BOHLANDER, S. K., HIEBERT, S. W., WARD, D. C., BRAY-WARD, P., MORGAN, E., RAIMONDI, S. C., ROWLEY, J. D. & GILLILAND, D. G. 1995. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A,* 92**,** 4917-21.

GOLUB, T. R., SLONIM, D. K., TAMAYO, P., HUARD, C., GAASENBEEK, M., MESIROV, J. P., COLLER, H., LOH, M. L., DOWNING, J. R., CALIGIURI, M. A., BLOOMFIELD, C. D. &

LANDER, E. S. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science,* 286**,** 531-7.

GOODE, D. K., OBIER, N., VIJAYABASKAR, M. S., LIE, A. L. M., LILLY, A. J., HANNAH, R., LICHTINGER, M., BATTA, K., FLORKOWSKA, M., PATEL, R., CHALLINOR, M., WALLACE, K., GILMOUR, J., ASSI, S. A., CAUCHY, P., HOOGENKAMP, M., WESTHEAD, D. R., LACAUD, G., KOUSKOFF, V., GOTTGENS, B. & BONIFER, C. 2016. Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev Cell,* 36**,** 572-87.

GOTTGENS, B. 2015. Regulatory network control of blood stem cells. *Blood,* 125**,** 2614-20.

GOYAMA, S., NITTA, E., YOSHINO, T., KAKO, S., WATANABE-OKOCHI, N., SHIMABE, M., IMAI, Y., TAKAHASHI, K. & KUROKAWA, M. 2010. EVI-1 interacts with histone methyltransferases SUV39H1 and G9a for transcriptional repression and bone marrow immortalization. *Leukemia,* 24**,** 81-88.

GOYAMA, S., SCHIBLER, J., CUNNINGHAM, L., ZHANG, Y., RAO, Y., NISHIMOTO, N., NAKAGAWA, M., OLSSON, A., WUNDERLICH, M., LINK, K. A., MIZUKAWA, B., GRIMES, H. L., KUROKAWA, M., LIU, P. P., HUANG, G. & MULLOY, J. C. 2013. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest,* 123**,** 3876-88.

GOYAMA, S., YAMAMOTO, G., SHIMABE, M., SATO, T., ICHIKAWA, M., OGAWA, S., CHIBA, S. & KUROKAWA, M. 2008. Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. *Cell Stem Cell,* 3**,** 207-220.

GREAVES, M. & MALEY, C. C. 2012. Clonal evolution in cancer. *Nature,* 481**,** 306-313.

GRIMWADE, D., HILLS, R. K., MOORMAN, A. V., WALKER, H., CHATTERS, S., GOLDSTONE, A. H., WHEATLEY, K., HARRISON, C. J., BURNETT, A. K. & NATIONAL CANCER RESEARCH INSTITUTE ADULT LEUKAEMIA WORKING, G. 2010. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood,* 116**,** 354-65.

GROSCHEL, S., LUGTHART, S., SCHLENK, R. F., VALK, P. J., EIWEN, K., GOUDSWAARD, C., VAN PUTTEN, W. J., KAYSER, S., VERDONCK, L. F., LUBBERT, M., OSSENKOPPELE, G. J., GERMING, U., SCHMIDT-WOLF, I., SCHLEGELBERGER, B., KRAUTER, J., GANSER, A., DOHNER, H., LOWENBERG, B., DOHNER, K. & DELWEL, R. 2010. High EVI1 expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. *J.Clin.Oncol.,* 28**,** 2101-2107.

GROSCHEL, S., SANDERS, M. A., HOOGENBOEZEM, R., DE WIT, E., BOUWMAN, B. A., ERPELINCK, C., VAN DER VELDEN, V. H., HAVERMANS, M., AVELLINO, R., VAN LOM, K., ROMBOUTS, E. J., VAN DUIN, M., DOHNER, K., BEVERLOO, H. B., BRADNER, J. E., DOHNER, H., LOWENBERG, B., VALK, P. J., BINDELS, E. M., DE LAAT, W. & DELWEL, R. 2014. A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. *Cell,* 157**,** 369-81.

GROSCHEL, S., SCHLENK, R. F., ENGELMANN, J., ROCKOVA, V., TELEANU, V., KUHN, M. W., EIWEN, K., ERPELINCK, C., HAVERMANS, M., LUBBERT, M., GERMING, U., SCHMIDT-WOLF, I. G., BEVERLOO, H. B., SCHUURHUIS, G. J., OSSENKOPPELE, G. J., SCHLEGELBERGER, B., VERDONCK, L. F., VELLENGA, E., VERHOEF, G., VANDENBERGHE, P., PABST, T., BARGETZI, M., KRAUTER, J., GANSER, A., VALK, P. J., LOWENBERG, B., DOHNER, K., DOHNER, H. & DELWEL, R. 2013. Deregulated expression of EVI1 defines a poor prognostic subset of MLL-rearranged acute myeloid leukemias: a study of the German-Austrian Acute Myeloid Leukemia Study Group and the Dutch-Belgian-Swiss HOVON/SAKK Cooperative Group. *J Clin Oncol,* 31**,** 95-103.

GROSVELD, F., VAN ASSENDELFT, G. B., GREAVES, D. R. & KOLLIAS, G. 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell,* 51**,** 975-985.

GROWNEY, J. D., SHIGEMATSU, H., LI, Z., LEE, B. H., ADELSPERGER, J., ROWAN, R., CURLEY, D. P., KUTOK, J. L., AKASHI, K., WILLIAMS, I. R., SPECK, N. A. & GILLILAND, D. G. 2005. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood,* 106**,** 494-504.

GU, T. L., GOETZ, T. L., GRAVES, B. J. & SPECK, N. A. 2000. Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1). *Mol Cell Biol,* 20**,** 91-103.

GUERMAH, M., PALHAN, V. B., TACKETT, A. J., CHAIT, B. T. & ROEDER, R. G. 2006. Synergistic Functions of SII and p300 in Productive Activator-Dependent Transcription of Chromatin Templates. *Cell,* 125**,** 275-286.

GUNTHER, C. V. & GRAVES, B. J. 1994. Identification of ETS domain proteins in murine T lymphocytes that interact with the Moloney murine leukemia virus enhancer. *Mol Cell Biol,* 14**,** 7569-80.

HARRIS, W. J., HUANG, X., LYNCH, J. T., SPENCER, G. J., HITCHIN, J. R., LI, Y., CICERI, F., BLASER, J. G., GREYSTOKE, B. F., JORDAN, A. M., MILLER, C. J., OGILVIE, D. J. & SOMERVAILLE, T. C. 2012. The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer cell,* 21**,** 473-87.

HARTMANN, L., DUTTA, S., OPATZ, S., VOSBERG, S., REITER, K., LEUBOLT, G., METZELER, K. H., HEROLD, T., BAMOPOULOS, S. A., BRAUNDL, K., ZELLMEIER, E., KSIENZYK, B., KONSTANDIN, N. P., SCHNEIDER, S., HOPFNER, K.-P., GRAF, A., KREBS, S., BLUM, H., MIDDEKE, J. M., STOLZEL, F., THIEDE, C., WOLF, S., BOHLANDER, S. K., PREISS, C., CHEN-WICHMANN, L., WICHMANN, C., SAUERLAND, M. C., BUCHNER, T., BERDEL, W. E., WORMANN, B. J., BRAESS, J., HIDDEMANN, W., SPIEKERMANN, K. & GREIF, P. A. 2016. ZBTB7A mutations in acute myeloid leukaemia with t(8;21) translocation. *Nat Commun,* 7.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell,* 38**,** 576-589.

HELBLING, D., MUELLER, B. U., TIMCHENKO, N. A., HAGEMEIJER, A., JOTTERAND, M., MEYER-MONARD, S., LISTER, A., ROWLEY, J. D., HUEGLI, B., FEY, M. F. & PABST, T. 2004. The leukemic fusion gene AML1-MDS1-EVI1 suppresses CEBPA in acute

myeloid leukemia by activation of Calreticulin. *Proc.Natl.Acad.Sci.U.S A,* 101**,** 13312-13317.

HERMANN, A., GOYAL, R. & JELTSCH, A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem,* 279**,** 48350-9.

HESSON, L. B., COOPER, W. N. & LATIF, F. 2007. Evaluation of the 3p21.3 tumour-suppressor gene cluster. *Oncogene,* 26**,** 7283-7301.

HIGUCHI, M., O'BRIEN, D., KUMARAVELU, P., LENNY, N., YEOH, E. J. & DOWNING, J. R. 2002. Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer cell,* 1**,** 63-74.

HILLS, R. K., RUSSELL, N., MILLIGAN, D., HUNTER, A. E., CLARK, R. E., BOWEN, D. & MCMULLIN, M. F. 2013. Reasons For Survival Improvement In Core Binding Factor AML: A 25 Year Analysis Of The UK MRC/NCRI AML Trials. *Blood,* 122**,** 358.

HIMES, S. R., TAGOH, H., GOONETILLEKE, N., SASMONO, T., OCEANDY, D., CLARK, R., BONIFER, C. & HUME, D. A. 2001. A highly conserved c-fms gene intronic element controls macrophage-specific and regulated expression. *J Leukoc Biol,* 70**,** 812-20.

HISA, T., SPENCE, S. E., RACHEL, R. A., FUJITA, M., NAKAMURA, T., WARD, J. M., DEVOR-HENNEMAN, D. E., SAIKI, Y., KUTSUNA, H., TESSAROLLO, L., JENKINS, N. A. & COPELAND, N. G. 2004. Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO J,* 23**,** 450-9.

HOCK, H., HAMBLEN, M. J., ROOKE, H. M., TRAVER, D., BRONSON, R. T., CAMERON, S. & ORKIN, S. H. 2003. Intrinsic Requirement for Zinc Finger Transcription Factor Gfi-1 in Neutrophil Differentiation. *Immunity,* 18**,** 109-120.

HOYT, P. R., C., B., DAVIS, A., YUTZEY, K., GAMER, L. W., POTTER, S. S., IHLE, J. & MUCENSKI, M. L. 1997. The Evi1 proto-oncogene is required at midgestation for neural, heart, and paraxial mesenchyme development. *Mech Dev,* 65**,** 55-70.

HU, Y., KIREEV, I., PLUTZ, M., ASHOURIAN, N. & BELMONT, A. S. 2009. Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *The Journal of Cell Biology,* 185**,** 87-100.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUANG, Y., SITWALA, K., BRONSTEIN, J., SANDERS, D., DANDEKAR, M., COLLINS, C., ROBERTSON, G., MACDONALD, J., CEZARD, T., BILENKY, M., THIESSEN, N., ZHAO, Y., ZENG, T., HIRST, M., HERO, A., JONES, S. & HESS, J. L. 2012. Identification and characterization of Hoxa9 binding sites in hematopoietic cells. *Blood,* 119**,** 388-398.

ICHIKAWA, M., ASAI, T., SAITO, T., YAMAMOTO, G., SEO, S., YAMAZAKI, I., YAMAGATA, T., MITANI, K., CHIBA, S., HIRAI, H., OGAWA, S. & KUROKAWA, M. 2004. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med,* 10**,** 299-304.

ILLENDULA, A., PULIKKAN, J. A., ZONG, H., GREMBECKA, J., XUE, L., SEN, S., ZHOU, Y., BOULTON, A., KUNTIMADDI, A., GAO, Y., RAJEWSKI, R. A., GUZMAN, M. L., CASTILLA, L. H. & BUSHWELLER, J. H. 2015. A small-molecule inhibitor of the aberrant transcription factor CBFβ-SMMHC delays leukemia in mice. *Science,* 347**,** 779-784.

ITO, T., KWON, H. Y., ZIMDAHL, B., CONGDON, K. L., BLUM, J., LENTO, W. E., ZHAO, C., LAGOO, A., GERRARD, G., FORONI, L., GOLDMAN, J., GOH, H., KIM, S.-H., KIM, D.-W., CHUAH, C., OEHLER, V. G., RADICH, J. P., JORDAN, C. T. & REYA, T. 2010. Regulation of myeloid leukaemia by the cell-fate determinant Musashi. *Nature,* 466**,** 765-768.

IVEY, A., HILLS, R. K., SIMPSON, M. A., JOVANOVIC, J. V., GILKES, A., GRECH, A., PATEL, Y., BHUDIA, N., FARAH, H., MASON, J., WALL, K., AKIKI, S., GRIFFITHS, M., SOLOMON, E., MCCAUGHAN, F., LINCH, D. C., GALE, R. E., VYAS, P., FREEMAN, S. D., RUSSELL, N., BURNETT, A. K. & GRIMWADE, D. 2016. Assessment of Minimal Residual Disease in Standard-Risk AML. *New England Journal of Medicine,* 374**,** 422-433.

IWASAKI, H. & AKASHI, K. 2007. Myeloid Lineage Commitment from the Hematopoietic Stem Cell. *Immunity,* 26**,** 726-740.

IWASAKI, H., MIZUNO, S.-I., WELLS, R. A., CANTOR, A. B., WATANABE, S. & AKASHI, K. 2003. GATA-1 Converts Lymphoid and Myelomonocytic Progenitors into the Megakaryocyte/Erythrocyte Lineages. *Immunity,* 19**,** 451-462.

IZUTSU, K., KUROKAWA, M., IMAI, Y., ICHIKAWA, M., ASAI, T., MAKI, K., MITANI, K. & HIRAI, H. 2002. The t(3;21) fusion product, AML1/Evi-1 blocks AML1-induced transactivation by recruiting CtBP. *Oncogene,* 21**,** 2695-2703.

IZUTSU, K., KUROKAWA, M., IMAI, Y., MAKI, K., MITANI, K. & HIRAI, H. 2001. The corepressor CtBP interacts with Evi-1 to repress transforming growth factor beta signaling. *Blood,* 97**,** 2815-2822.

JAISWAL , S., FONTANILLAS , P., FLANNICK , J., MANNING , A., GRAUMAN , P. V., MAR , B. G., LINDSLEY , R. C., MERMEL , C. H., BURTT , N., CHAVEZ , A., HIGGINS , J. M., MOLTCHANOV , V., KUO , F. C., KLUK , M. J., HENDERSON , B., KINNUNEN , L., KOISTINEN , H. A., LADENVALL , C., GETZ , G., CORREA , A., BANAHAN , B. F., GABRIEL , S., KATHIRESAN , S., STRINGHAM , H. M., MCCARTHY , M. I., BOEHNKE , M., TUOMILEHTO , J., HAIMAN , C., GROOP , L., ATZMON , G., WILSON , J. G., NEUBERG , D., ALTSHULER , D. & EBERT , B. L. 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine,* 371**,** 2488-2498.

JAISWAL, S., JAMIESON, C. H. M., PANG, W. W., PARK, C. Y., CHAO, M. P., MAJETI, R., TRAVER, D., VAN ROOIJEN, N. & WEISSMAN, I. L. 2009. CD47 Is Upregulated on Circulating Hematopoietic Stem Cells and Leukemia Cells to Avoid Phagocytosis. *Cell,* 138**,** 271-285.

JAMIESON, C. H., AILLES, L. E., DYLLA, S. J., MUIJTJENS, M., JONES, C., ZEHNDER, J. L., GOTLIB, J., LI, K., MANZ, M. G., KEATING, A., SAWYERS, C. L. & WEISSMAN, I. L. 2004. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med,* 351**,** 657-67.

JANICKI, S. M., TSUKAMOTO, T., SALGHETTI, S. E., TANSEY, W. P., SACHIDANANDAM, R., PRASANTH, K. V., RIED, T., SHAV-TAL, Y., BERTRAND, E., SINGER, R. H. & SPECTOR, D. L. 2004. From Silencing to Gene Expression: Real-Time Analysis in Single Cells. *Cell,* 116**,** 683-698.

JIN, G., YAMAZAKI, Y., TAKUWA, M., TAKAHARA, T., KANEKO, K., KUWATA, T., MIYATA, S. & NAKAMURA, T. 2007. Trib1 and Evi1 cooperate with Hoxa and Meis1 in myeloid leukemogenesis. *Blood,* 109**,** 3998-4005.

JONKERS, I. & LIS, J. T. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol,* 16**,** 167-77.

KAMEI, Y., XU, L., HEINZEL, T., TORCHIA, J., KUROKAWA, R., GLOSS, B., LIN, S. C., HEYMAN, R. A., ROSE, D. W., GLASS, C. K. & ROSENFELD, M. G. 1996. A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell,* 85**,** 403-14.

KATAOKA, K., SATO, T., YOSHIMI, A., GOYAMA, S., TSURUTA, T., KOBAYASHI, H., SHIMABE, M., ARAI, S., NAKAGAWA, M., IMAI, Y., KUMANO, K., KUMAGAI, K., KUBOTA, N., KADOWAKI, T. & KUROKAWA, M. 2011. Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. *J.Exp.Med.,* 208**,** 2403-2416.

KAWAGOE, H., KANDILCI, A., KRANENBURG, T. A. & GROSVELD, G. C. 2007. Overexpression of N-Myc rapidly causes acute myeloid leukemia in mice. *Cancer Res,* 67**,** 10677-85.

KELLY, T. K., DE CARVALHO, D. D. & JONES, P. A. 2010. Epigenetic modifications as therapeutic targets. *Nat Biotech,* 28**,** 1069-1078.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KIKUSHIGE, Y., MIYAMOTO, T., YUDA, J., JABBARZADEH-TABRIZI, S., SHIMA, T., TAKAYANAGI, S.-I., NIIRO, H., YURINO, A., MIYAWAKI, K., TAKENAKA, K., IWASAKI, H. & AKASHI, K. 2011. A TIM-3/Gal-9 Autocrine Stimulatory Loop Drives Self-Renewal of Human Myeloid Leukemia Stem Cells and Leukemic Progression. *Cell Stem Cell,* 17**,** 341-352.

KIM, E., ILAGAN, JANINE O., LIANG, Y., DAUBNER, GERRIT M., LEE, STANLEY C. W., RAMAKRISHNAN, A., LI, Y., CHUNG, YOUNG R., MICOL, J.-B., MURPHY, M. E., CHO, H., KIM, M.-K., ZEBARI, AHMAD S., AUMANN, S., PARK, CHRISTOPHER Y., BUONAMICI, S., SMITH, PETER G., DEEG, H. J., LOBRY, C., AIFANTIS, I., MODIS, Y., ALLAIN, FREDERIC H. T., HALENE, S., BRADLEY, ROBERT K. & ABDEL-WAHAB, O. 2015. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer cell,* 27**,** 617-630.

KIM, H. J., AHN, H. K., JUNG, C. W., MOON, J. H., PARK, C. H., LEE, K. O., KIM, S. H., KIM, Y. K., SOHN, S. K., LEE, W. S., KIM, K. H., MUN, Y. C., KIM, H., PARK, J., MIN, W. S. & KIM, D. H. 2013. KIT D816 mutation associates with adverse outcomes in core binding factor acute myeloid leukemia, especially in the subgroup with RUNX1/RUNX1T1 rearrangement. *Ann Hematol,* 92**,** 163-71.

KIREEVA, M. L., HANCOCK, B., CREMONA, G. H., WALTER, W., STUDITSKY, V. M. & KASHLEV, M. 2005. Nature of the nucleosomal barrier to RNA polymerase II. *Molecular Cell,* 18**,** 97-108.

KIREEVA, M. L., WALTER, W., TCHERNAJENKO, V., BONDARENKO, V., KASHLEV, M. & STUDITSKY, V. M. 2002. Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription. *Molecular Cell,* 9**,** 541-552.

KITABAYASHI, I., YOKOYAMA, A., SHIMIZU, K. & OHKI, M. 1998. Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J,* 17**,** 2994-3004.

KLCO, J. M., SPENCER, D. H., LAMPRECHT, T. L., SARKARIA, S. M., WYLIE, T., MAGRINI, V., HUNDAL, J., WALKER, J., VARGHESE, N., ERDMANN-GILMORE, P., LICHTI, C. F., MEYER, M. R., TOWNSEND, R. R., WILSON, R. K., MARDIS, E. R. & LEY, T. J. 2013. Genomic impact of transient low-dose decitabine treatment on primary AML cells. *Blood,* 121**,** 1633-43.

KLEMSZ, M. J., MCKERCHER, S. R., CELADA, A., VAN BEVEREN, C. & MAKI, R. A. 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell,* 61**,** 113-124.

KLESNEY-TAIT, J., KECK, K., LI, X., GILFILLAN, S., OTERO, K., BARUAH, S., MEYERHOLZ, D. K., VARGA, S. M., KNUDSON, C. J., MONINGER, T. O., MORELAND, J., ZABNER, J. & COLONNA, M. 2013. Transepithelial migration of neutrophils into the lung requires TREM-1. *J Clin Invest,* 123**,** 138-49.

KOMINE, O., HAYASHI, K., NATSUME, W., WATANABE, T., SEKI, Y., SEKI, N., YAGI, R., SUKZUKI, W., TAMAUCHI, H., HOZUMI, K., HABU, S., KUBO, M. & SATAKE, M. 2003. The Runx1 Transcription Factor Inhibits the Differentiation of Naive CD4+ T Cells into the Th2 Lineage by Repressing GATA3 Expression. *The Journal of Experimental Medicine,* 198**,** 51-61.

KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell,* 91**,** 661-672.

KRAUTH, M. T., EDER, C., ALPERMANN, T., BACHER, U., NADARAJAH, N., KERN, W., HAFERLACH, C., HAFERLACH, T. & SCHNITTGER, S. 2014. High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome. *Leukemia*.

KREIDER, B. L., ORKIN, S. H. & IHLE, J. N. 1993. Loss of erythropoietin responsiveness in erythroid progenitors due to expression of the Evi-1 myeloid-transforming gene. *Proc.Natl.Acad.Sci.U.S A,* 90**,** 6454-6458.

KRIVTSOV, A. V., FIGUEROA, M. E., SINHA, A. U., STUBBS, M. C., FENG, Z., VALK, P. J., DELWEL, R., DOHNER, K., BULLINGER, L., KUNG, A. L., MELNICK, A. M. & ARMSTRONG, S. A. 2013. Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia,* 27**,** 852-60.

KRIVTSOV, A. V., TWOMEY, D., FENG, Z., STUBBS, M. C., WANG, Y., FABER, J., LEVINE, J. E., WANG, J., HAHN, W. C., GILLILAND, D. G., GOLUB, T. R. & ARMSTRONG, S. A. 2006. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature,* 442**,** 818-22.

KROON, E., KROSL, J., THORSTEINSDOTTIR, U., BABAN, S., BUCHBERG, A. M. & SAUVAGEAU, G. 1998. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *EMBO J,* 17**,** 3714-25.

KRYLOV, D., OLIVE, M. & VINSON, C. 1995. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J,* 14**,** 5329-37.

KUROKAWA, M., MITANI, K., IMAI, Y., OGAWA, S., YAZAKI, Y. & HIRAI, H. 1998a. The t(3;21) fusion product, AML1/Evi-1, interacts with Smad3 and blocks transforming growth factor-beta-mediated growth inhibition of myeloid cells. *Blood,* 92**,** 4003-4012.

KUROKAWA, M., MITANI, K., IRIE, K., MATSUYAMA, T., TAKAHASHI, T., CHIBA, S., YAZAKI, Y., MATSUMOTO, K. & HIRAI, H. 1998b. The oncoprotein Evi-1 represses TGF-beta signalling by inhibiting Smad3. *Nature,* 394**,** 92-96.

KWASNIESKI, J. C., FIORE, C., CHAUDHARI, H. G. & COHEN, B. A. 2014. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res,* 24**,** 1595-1602.

KWOK, R. P., LUNDBLAD, J. R., CHRIVIA, J. C., RICHARDS, J. P., BACHINGER, H. P., BRENNAN, R. G., ROBERTS, S. G., GREEN, M. R. & GOODMAN, R. H. 1994. Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature,* 370**,** 223-6.

LAIOSA, C. V., STADTFELD, M., XIE, H., DE ANDRES-AGUAYO, L. & GRAF, T. 2006. Reprogramming of Committed T Cell Progenitors to Macrophages and Dendritic Cells by C/EBPα and PU.1 Transcription Factors. *Immunity,* 25**,** 731-744.

LAPIDOT, T., SIRARD, C., VORMOOR, J., MURDOCH, B., HOANG, T., CACERES-CORTES, J., MINDEN, M., PATERSON, B., CALIGIURI, M. A. & DICK, J. E. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature,* 367**,** 645-8.

LARICCHIA-ROBBIO, L., FAZZINA, R., LI, D., RINALDI, C. R., SINHA, K. K., CHAKRABORTY, S. & NUCIFORA, G. 2006. Point mutations in two EVI1 Zn fingers abolish EVI1-GATA1 interaction and allow erythroid differentiation of murine bone marrow cells. *Mol Cell Biol,* 26**,** 7658-66.

LARICCHIA-ROBBIO, L., PREMANAND, K., RINALDI, C. R. & NUCIFORA, G. 2009. EVI1 Impairs myelopoiesis by deregulation of PU.1 function. *Cancer Res.,* 69**,** 1633-1642.

LASSAR, A. B., DAVIS, R. L., WRIGHT, W. E., KADESCH, T., MURRE, C., VORONOVA, A., BALTIMORE, D. & WEINTRAUB, H. 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell,* 66**,** 305-15.

LAVALLÉE, V.-P., LEMIEUX, S., BOUCHER, G., GENDRON, P., BOIVIN, I., ARMSTRONG, R. N., SAUVAGEAU, G. & HÉBERT, J. 2016. RNA-sequencing analysis of core binding factor AML identifies recurrent ZBTB7A mutations and defines RUNX1-CBFA2T3 fusion signature. *Blood,* 127**,** 2498-2501.

LAWRENCE, H. J., CHRISTENSEN, J., FONG, S., HU, Y.-L., WEISSMAN, I., SAUVAGEAU, G., HUMPHRIES, R. K. & LARGMAN, C. 2005. Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. *Blood,* 106**,** 3988-3994.

LAWRENCE, H. J., ROZENFELD, S., CRUZ, C., MATSUKUMA, K., KWONG, A., KOMUVES, L., BUCHBERG, A. M. & LARGMAN, C. 1999. Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. *Leukemia,* 13**,** 1993-9.

LEDDIN, M., PERROD, C., HOOGENKAMP, M., GHANI, S., ASSI, S., HEINZ, S., WILSON, N. K., FOLLOWS, G., SCHÖNHEIT, J., VOCKENTANZ, L., MOSAMMAM, A. M., CHEN, W., TENEN, D. G., WESTHEAD, D. R., GÖTTGENS, B., BONIFER, C. & ROSENBAUER, F. 2011. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood,* 117**,** 2827-2838.

LEGRAVEREND, C., ANTONSON, P., FLODBY, P. & XANTHOPOULOS, K. G. 1993. High level activity of the mouse CCAAT/enhancer binding protein (C/EBP alpha) gene promoter involves autoregulation and several ubiquitous transcription factors. *Nucleic Acids Res,* 21**,** 1735-42.

LI, B., CAREY, M. & WORKMAN, J. L. 2007. The Role of Chromatin during Transcription. *Cell,* 128**,** 707-719.

LI, Y., OKUNO, Y., ZHANG, P., RADOMSKA, H. S., CHEN, H.-M., IWASAKI, H., AKASHI, K., KLEMSZ, M. J., MCKERCHER, S. R., MAKI, R. A. & TENEN, D. G. 2001. Regulation of the PU.1 gene by distal elements. *Blood,* 98**,** 2958-2965.

LING, K.-W., OTTERSBACH, K., VAN HAMBURG, J. P., OZIEMLAK, A., TSAI, F.-Y., ORKIN, S. H., PLOEMACHER, R., HENDRIKS, R. W. & DZIERZAK, E. 2004. GATA-2 Plays Two Functionally Distinct Roles during the Ontogeny of Hematopoietic Stem Cells. *The Journal of Experimental Medicine,* 200**,** 871-882.

LIU, P., TARLE, S., HAJRA, A., CLAXTON, D., MARLTON, P., FREEDMAN, M., SICILIANO, M. & COLLINS, F. 1993. Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science,* 261**,** 1041-1044.

LIU, X.-S., HAINES, J. E., MEHANNA, E. K., GENET, M. D., BEN-SAHRA, I., ASARA, J. M., MANNING, B. D. & YUAN, Z.-M. 2014. ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis. *Genes & Development,* 28**,** 1917-1928.

LIU, Y., CHEN, W., GAUDET, J., CHENEY, M. D., ROUDAIA, L., CIERPICKI, T., KLET, R. C., HARTMAN, K., LAUE, T. M., SPECK, N. A. & BUSHWELLER, J. H. 2007. Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity. *Cancer cell,* 11**,** 483-97.

LIU, Y., CHENEY, M. D., GAUDET, J. J., CHRUSZCZ, M., LUKASIK, S. M., SUGIYAMA, D., LARY, J., COLE, J., DAUTER, Z., MINOR, W., SPECK, N. A. & BUSHWELLER, J. H. 2006. The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer cell,* 9**,** 249-60.

LOKE, J., KHAN, J. N., WILSON, J. S., CRADDOCK, C. & WHEATLEY, K. 2015. Mylotarg has potent anti-leukaemic effect: a systematic review and meta-analysis of anti-CD33 antibody treatment in acute myeloid leukaemia. *Ann Hematol,* 94**,** 361-73.

LOUGHRAN, S. J., KRUSE, E. A., HACKING, D. F., DE GRAAF, C. A., HYLAND, C. D., WILLSON, T. A., HENLEY, K. J., ELLIS, S., VOSS, A. K., METCALF, D., HILTON, D. J., ALEXANDER, W. S. & KILE, B. T. 2008. The transcription factor Erg is essential for

definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol,* 9**,** 810-9.

LOZZIO, C. & LOZZIO, B. 1975. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood,* 45**,** 321-334.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.

LUGTHART, S., GROSCHEL, S., BEVERLOO, H. B., KAYSER, S., VALK, P. J., VAN ZELDEREN-BHOLA, S. L., JAN, O. G., VELLENGA, E., RUITER, V. D. B.-D., SCHANZ, U., VERHOEF, G., VANDENBERGHE, P., FERRANT, A., KOHNE, C. H., PFREUNDSCHUH, M., HORST, H. A., KOLLER, E., VON LILIENFELD-TOAL, M., BENTZ, M., GANSER, A., SCHLEGELBERGER, B., JOTTERAND, M., KRAUTER, J., PABST, T., THEOBALD, M., SCHLENK, R. F., DELWEL, R., DOHNER, K., LOWENBERG, B. & DOHNER, H. 2010. Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. *J.Clin.Oncol.,* 28**,** 3890-3898.

LUNARDI, A., GUARNERIO, J., WANG, G., MAEDA, T. & PANDOLFI, P. P. 2013. Role of LRF/Pokemon in lineage fate decisions. *Blood,* 121**,** 2845-2853.

LUO, B. & LEE, A. S. 2013. The critical roles of endoplasmic reticulum chaperones and unfolded protein response in tumorigenesis and anticancer therapies. *Oncogene,* 32**,** 805-18.

LUTTERBACH, B. & HIEBERT, S. W. 2000. Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene,* 245**,** 223-235.

LUTTERBACH, B., WESTENDORF, J. J., LINGGI, B., PATTEN, A., MONIWA, M., DAVIE, J. R., HUYNH, K. D., BARDWELL, V. J., LAVINSKY, R. M., ROSENFELD, M. G., GLASS, C., SETO, E. & HIEBERT, S. W. 1998. ETO, a Target of t(8;21) in Acute Leukemia, Interacts with the N-CoR and mSin3 Corepressors. *Molecular and Cellular Biology,* 18**,** 7176-7184.

MAERE, S., HEYMANS, K. & KUIPER, M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics,* 21**,** 3448-9.

MAKI, K., YAMAGATA, T., ASAI, T., YAMAZAKI, I., ODA, H., HIRAI, H. & MITANI, K. 2005. Dysplastic definitive hematopoiesis in AML1/EVI1 knock-in embryos. *Blood,* 106**,** 2147-2155.

MAKI, K., YAMAGATA, T., YAMAZAKI, I., ODA, H. & MITANI, K. 2006. Development of megakaryoblastic leukaemia in Runx1-Evi1 knock-in chimaeric mouse. *Leukemia,* 20**,** 1458-1460.

MARTENS, J. H., MANDOLI, A., SIMMER, F., WIERENGA, B. J., SAEED, S., SINGH, A. A., ALTUCCI, L., VELLENGA, E. & STUNNENBERG, H. G. 2012. ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood,* 120**,** 4038-48.

MASUDA, T., WANG, X., MAEDA, M., CANVER, M. C., SHER, F., FUNNELL, A. P., FISHER, C., SUCIU, M., MARTYN, G. E., NORTON, L. J., ZHU, C., KURITA, R., NAKAMURA, Y., XU, J., HIGGS, D. R., CROSSLEY, M., BAUER, D. E., ORKIN, S. H., KHARCHENKO,

P. V. & MAEDA, T. 2016. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science,* 351**,** 285-9.

MATHENY, C. J., SPECK, M. E., CUSHING, P. R., ZHOU, Y., CORPORA, T., REGAN, M., NEWMAN, M., ROUDAIA, L., SPECK, C. L., GU, T. L., GRIFFEY, S. M., BUSHWELLER, J. H. & SPECK, N. A. 2007. Disease mutations in RUNX1 and RUNX2 create nonfunctional, dominant-negative, or hypomorphic alleles. *EMBO J,* 26**,** 1163-75.

MAZUMDAR, C., SHEN, Y., XAVY, S., ZHAO, F., REINISCH, A., LI, R., CORCES, M. R., FLYNN, RYAN A., BUENROSTRO, JASON D., CHAN, STEVEN M., THOMAS, D., KOENIG, JULIE L., HONG, W.-J., CHANG, HOWARD Y. & MAJETI, R. 2015. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. *Cell Stem Cell,* 17**,** 675-688.

MEDVINSKY, A. & DZIERZAK, E. 1996. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell,* 86**,** 897-906.

MEDVINSKY, A. L., SAMOYLINA, N. L., MULLER, A. M. & DZIERZAK, E. A. 1993. An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature,* 364**,** 64-7.

MERCER, E. M., LIN, Y. C. & MURRE, C. 2011. Factors and networks that underpin early hematopoiesis. *Semin Immunol,* 23**,** 317-25.

METZELER, K. H., DUFOUR, A., BENTHAUS, T., HUMMEL, M., SAUERLAND, M. C., HEINECKE, A., BERDEL, W. E., BUCHNER, T., WORMANN, B., MANSMANN, U., BRAESS, J., SPIEKERMANN, K., HIDDEMANN, W., BUSKE, C. & BOHLANDER, S. K. 2009. ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: a comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. *J Clin Oncol,* 27**,** 5031-8.

MEYER, S. E., QIN, T., MUENCH, D. E., MASUDA, K., VENKATASUBRAMANIAN, M., ORR, E., SUAREZ, L., GORE, S. D., DELWEL, R., PAIETTA, E., TALLMAN, M. S., FERNANDEZ, H., MELNICK, A., LE BEAU, M. M., KOGAN, S., SALOMONIS, N., FIGUEROA, M. E. & GRIMES, H. L. 2016. Dnmt3a haploinsufficiency transforms Flt3-ITD myeloproliferative disease into a rapid, spontaneous, and fully-penetrant acute myeloid leukemia. *Cancer Discovery*.

MEYERS, S., DOWNING, J. R. & HIEBERT, S. W. 1993. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein-protein interactions. *Molecular and Cellular Biology,* 13**,** 6336-6345.

MIKKOLA, H. K. A., KLINTMAN, J., YANG, H., HOCK, H., SCHLAEGER, T. M., FUJIWARA, Y. & ORKIN, S. H. 2003. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature,* 421**,** 547-551.

MITANI, K., OGAWA, S., TANAKA, T., MIYOSHI, H., KUROKAWA, M., MANO, H., YAZAKI, Y., OHKI, M. & HIRAI, H. 1994. Generation of the AML1-EVI-1 fusion gene in the t(3;21)(q26;q22) causes blastic crisis in chronic myelocytic leukemia. *EMBO J,* 13**,** 504-510.

MIYAMOTO, T., WEISSMAN, I. L. & AKASHI, K. 2000. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proceedings of the National Academy of Sciences,* 97**,** 7521-7526.

MIYOSHI, H., KOZU, T., SHIMIZU, K., ENOMOTO, K., MASEKI, N., KANEKO, Y., KAMADA, N. & OHKI, M. 1993. The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. *EMBO J,* 12**,** 2715-21.

MORISHITA, K., SUZUKAWA, K., TAKI, T., IHLE, J. N. & YOKOTA, J. 1995. EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus sequence of GACAAGATAAGATAAN1-28 CTCATCTTC. *Oncogene,* 10**,** 1961-7.

MORRISON, S. J., HEMMATI, H. D., WANDYCZ, A. M. & WEISSMAN, I. L. 1995. The purification and characterization of fetal liver hematopoietic stem cells. *Proc Natl Acad Sci U S A,* 92**,** 10302-6.

MUCENSKI, M. L., TAYLOR, B. A., IHLE, J. N., HARTLEY, J. W., MORSE, H. C., III, JENKINS, N. A. & COPELAND, N. G. 1988. Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Mol.Cell Biol.,* 8**,** 301-308.

MULLER-SIEBURG, C. E., WHITLOCK, C. A. & WEISSMAN, I. L. 1986. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1-lo hematopoietic stem cell. *Cell,* 44**,** 653-62.

MULLER, A. M., MEDVINSKY, A., STROUBOULIS, J., GROSVELD, F. & DZIERZAK, E. 1994. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity,* 1**,** 291-301.

MULLOY, J. C., CAMMENGA, J., MACKENZIE, K. L., BERGUIDO, F. J., MOORE, M. A. S. & NIMER, S. D. 2002. The AML1-ETO fusion protein promotes the expansion of human hematopoietic stem cells. *Blood,* 99**,** 15-23.

MUPO, A., CELANI, L., DOVEY, O., COOPER, J. L., GROVE, C., RAD, R., SPORTOLETTI, P., FALINI, B., BRADLEY, A. & VASSILIOU, G. S. 2013. A powerful molecular synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in mice. *Leukemia,* 27**,** 1917-1920.

NAN, X., CAMPOY, F. J. & BIRD, A. 1997. MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell,* 88**,** 471-481.

NAN, X., NG, H. H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R. N. & BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-389.

NEPH, S., VIERSTRA, J., STERGACHIS, A. B., REYNOLDS, A. P., HAUGEN, E., VERNOT, B., THURMAN, R. E., JOHN, S., SANDSTROM, R., JOHNSON, A. K., MAURANO, M. T., HUMBERT, R., RYNES, E., WANG, H., VONG, S., LEE, K., BATES, D., DIEGEL, M., ROACH, V., DUNN, D., NERI, J., SCHAFER, A., HANSEN, R. S., KUTYAVIN, T., GISTE, E., WEAVER, M., CANFIELD, T., SABO, P., ZHANG, M., BALASUNDARAM, G., BYRON, R., MACCOSS, M. J., AKEY, J. M., BENDER, M. A., GROUDINE, M., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature,* 489**,** 83-90.

NICK, H. J., KIM, H. G., CHANG, C. W., HARRIS, K. W., REDDY, V. & KLUG, C. A. 2012. Distinct classes of c-Kit-activating mutations differ in their ability to promote RUNX1-ETO-associated acute myeloid leukemia. *Blood,* 119**,** 1522-31.

NORTH, T. E., DE BRUIJN, M. F., STACY, T., TALEBIAN, L., LIND, E., ROBIN, C., BINDER, M., DZIERZAK, E. & SPECK, N. A. 2002. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity,* 16**,** 661-72.

NOVERSHTERN, N., SUBRAMANIAN, A., LAWTON, L. N., MAK, R. H., HAINING, W. N., MCCONKEY, M. E., HABIB, N., YOSEF, N., CHANG, C. Y., SHAY, T., FRAMPTON, G. M., DRAKE, A. C. B., LESKOV, I., NILSSON, B., PREFFER, F., DOMBKOWSKI, D., EVANS, J. W., LIEFELD, T., SMUTKO, J. S., CHEN, J., FRIEDMAN, N., YOUNG, R. A., GOLUB, T. R., REGEV, A. & EBERT, B. L. 2011. Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. *Cell,* 144**,** 296-309.

NUCIFORA, G., BEGY, C. R., KOBAYASHI, H., ROULSTON, D., CLAXTON, D., PEDERSEN-BJERGAARD, J., PARGANAS, E., IHLE, J. N. & ROWLEY, J. D. 1994. Consistent intergenic splicing and production of multiple transcripts between AML1 at 21q22 and unrelated genes at 3q26 in (3;21)(q26;q22) translocations. *Proc.Natl.Acad.Sci.U.S A,* 91**,** 4004-4008.

NUKINA, A., KAGOYA, Y., WATANABE-OKOCHI, N., ARAI, S., UEDA, K., YOSHIMI, A., NANNYA, Y. & KUROKAWA, M. 2014. Single-cell gene expression analysis reveals clonal architecture of blast-phase chronic myeloid leukaemia. *British Journal of Haematology,* 165**,** 414-416.

NUTT, S. L., HEAVEY, B., ROLINK, A. G. & BUSSLINGER, M. 1999. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature,* 401**,** 556-562.

OELGESCHLAGER, M., JANKNECHT, R., KRIEG, J., SCHREEK, S. & LUSCHER, B. 1996. Interaction of the co-activator CBP with Myb proteins: effects on Myb-specific transactivation and on the cooperativity with NF-M. *EMBO J,* 15**,** 2771-80.

OHLSSON, E., HASEMANN, M. S., WILLER, A., LAURIDSEN, F. K. B., RAPIN, N., JENDHOLM, J. & PORSE, B. T. 2014. Initiation of MLL-rearranged AML is dependent on C/EBPα. *The Journal of Experimental Medicine,* 211**,** 5-13.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-257.

OKUDA, T., CAI, Z., YANG, S., LENNY, N., LYU, C. J., VAN DEURSEN, J. M. A., HARADA, H. & DOWNING, J. R. 1998. Expression of a Knocked-In AML1-ETO Leukemia Gene Inhibits the Establishment of Normal Definitive Hematopoiesis and Directly Generates Dysplastic Hematopoietic Progenitors. *Blood,* 91**,** 3134-3143.

OKUDA, T., VAN DEURSEN, J., HIEBERT, S. W., GROSVELD, G. & DOWNING, J. R. 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell,* 84**,** 321-30.

OSATO, M., ASOU, N., ABDALLA, E., HOSHINO, K., YAMASAKI, H., OKUBO, T., SUZUSHIMA, H., TAKATSUKI, K., KANNO, T., SHIGESADA, K. & ITO, Y. 1999. Biallelic and Heterozygous Point Mutations in the Runt Domain of

theAML1/PEBP2B Gene Associated With Myeloblastic Leukemias. *Blood,* 93**,** 1817-1824.

OWEN-HUGHES, T., UTLEY, R. T., XF, XE, J., PETERSON, C. L. & WORKMAN, J. L. 1996. Persisent Site-Specific Remodeling of a Nucleosome Array by Transient Action of the SWI/SNF Complex. *Science,* 273**,** 513-516.

PABST, T., MUELLER, B. U., HARAKAWA, N., SCHOCH, C., HAFERLACH, T., BEHRE, G., HIDDEMANN, W., ZHANG, D. E. & TENEN, D. G. 2001. AML1-ETO downregulates the granulocytic differentiation factor C/EBP[alpha] in t(8;21) myeloid leukemia. *Nat Med,* 7**,** 444-451.

PALMER, S., BROUILLET, J. P., KILBEY, A., FULTON, R., WALKER, M., CROSSLEY, M. & BARTHOLOMEW, C. 2001. Evi-1 transforming and repressor activities are mediated by CtBP co-repressor proteins. *J.Biol.Chem.,* 276**,** 25834-25840.

PAN, X., MINEGISHI, N., HARIGAE, H., YAMAGIWA, H., MINEGISHI, M., AKINE, Y. & YAMAMOTO, M. 2000. Identification of Human GATA-2 Gene Distal IS Exon and Its Expression in Hematopoietic Stem Cell Fractions. *Journal of Biochemistry,* 127**,** 105-112.

PAPAEMMANUIL, E., GERSTUNG, M., BULLINGER, L., GAIDZIK, V. I., PASCHKA, P., ROBERTS, N. D., POTTER, N. E., HEUSER, M., THOL, F., BOLLI, N., GUNDEM, G., VAN LOO, P., MARTINCORENA, I., GANLY, P., MUDIE, L., MCLAREN, S., O'MEARA, S., RAINE, K., JONES, D. R., TEAGUE, J. W., BUTLER, A. P., GREAVES, M. F., GANSER, A., DÖHNER, K., SCHLENK, R. F., DÖHNER, H. & CAMPBELL, P. J. 2016. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *New England Journal of Medicine,* 374**,** 2209-2221.

PAQUETTE, R. L., NICOLL, J., CHALUKYA, M., ELASHOFF, D., SHAH, N. P., SAWYERS, C., SPITERI, E., NANJANGUD, G. & RAO, P. N. 2011. Frequent EVI1 translocations in myeloid blast crisis CML that evolves through tyrosine kinase inhibitors. *Cancer Genet,* 204**,** 392-7.

PARK, S. H., CHI, H. S., MIN, S. K., PARK, B. G., JANG, S. & PARK, C. J. 2011. Prognostic impact of c-KIT mutations in core binding factor acute myeloid leukemia. *Leuk Res,* 35**,** 1376-83.

PASCHKA, P. & DOHNER, K. 2013. Core-binding factor acute myeloid leukemia: can we improve on HiDAC consolidation? *Hematology Am Soc Hematol Educ Program,* 2013**,** 209-19.

PASSEGUE, E., JAMIESON, C. H., AILLES, L. E. & WEISSMAN, I. L. 2003. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci U S A,* 100 Suppl 1**,** 11842-9.

PAUL, F., ARKIN, Y. A., GILADI, A., JAITIN, DIEGO A., KENIGSBERG, E., KEREN-SHAUL, H., WINTER, D., LARA-ASTIASO, D., GURY, M., WEINER, A., DAVID, E., COHEN, N., LAURIDSEN, FELICIA KATHRINE B., HAAS, S., SCHLITZER, A., MILDNER, A., GINHOUX, F., JUNG, S., TRUMPP, A., PORSE, BO T., TANAY, A. & AMIT, I. 2015. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell,* 163**,** 1663-1677.

PEDERSEN, T. A., KOWENZ-LEUTZ, E., LEUTZ, A. & NERLOV, C. 2001. Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes Dev,* 15**,** 3208-16.

PERKINS, A. S., FISHEL, R., JENKINS, N. A. & COPELAND, N. G. 1991. Evi-1, a murine zinc finger proto-oncogene, encodes a sequence-specific DNA-binding protein. *Mol Cell Biol,* 11**,** 2665-74.

PERSONS, D. A., ALLAY, J. A., ALLAY, E. R., ASHMUN, R. A., ORLIC, D., JANE, S. M., CUNNINGHAM, J. M. & NIENHUIS, A. W. 1999. Enforced Expression of the GATA-2 Transcription Factor Blocks Normal Hematopoiesis. *Blood,* 93**,** 488-499.

PETROVICK, M. S., HIEBERT, S. W., FRIEDMAN, A. D., HETHERINGTON, C. J., TENEN, D. G. & ZHANG, D. E. 1998. Multiple functional domains of AML1: PU.1 and C/EBPalpha synergize with different regions of AML1. *Mol Cell Biol,* 18**,** 3915-25.

PHILLIPS, R. L., ERNST, R. E., BRUNK, B., IVANOVA, N., MAHAN, M. A., DEANEHAN, J. K., MOORE, K. A., OVERTON, G. C. & LEMISCHKA, I. R. 2000. The genetic program of hematopoietic stem cells. *Science,* 288**,** 1635-40.

PIMANDA, J. E. & GOTTGENS, B. 2010. Gene regulatory networks governing haematopoietic stem cell development and identity. *Int J Dev Biol,* 54**,** 1201-11.

PIMANDA, J. E., OTTERSBACH, K., KNEZEVIC, K., KINSTON, S., CHAN, W. Y. I., WILSON, N. K., LANDRY, J.-R., WOOD, A. D., KOLB-KOKOCINSKI, A., GREEN, A. R., TANNAHILL, D., LACAUD, G., KOUSKOFF, V. & GÖTTGENS, B. 2007. Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proceedings of the National Academy of Sciences,* 104**,** 17692-17697.

PINHEIRO, I., MARGUERON, R., SHUKEIR, N., EISOLD, M., FRITZSCH, C., RICHTER, F., MITTLER, G., GENOUD, C., GOYAMA, S., KUROKAWA, M., SON, J., REINBERG, D., LACHNER, M. & JENUWEIN, T. 2012. Prdm3 and Prdm16 are H3K9me1 Methyltransferases Required for Mammalian Heterochromatin Integrity. *Cell,* 150**,** 948-960.

PIPER, J., ELZE, M. C., CAUCHY, P., COCKERILL, P. N., BONIFER, C. & OTT, S. 2013. Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res,* 41**,** e201.

POLLARD, J. A., ALONZO, T. A., GERBING, R. B., HO, P. A., ZENG, R., RAVINDRANATH, Y., DAHL, G., LACAYO, N. J., BECTON, D., CHANG, M., WEINSTEIN, H. J., HIRSCH, B., RAIMONDI, S. C., HEEREMA, N. A., WOODS, W. G., LANGE, B. J., HURWITZ, C., ARCECI, R. J., RADICH, J. P., BERNSTEIN, I. D., HEINRICH, M. C. & MESHINCHI, S. 2010. Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. *Blood,* 115**,** 2372-9.

PORCHER, C., LIAO, E. C., FUJIWARA, Y., ZON, L. I. & ORKIN, S. H. 1999. Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development,* 126**,** 4603-4615.

PORCHER, C., SWAT, W., ROCKWELL, K., FUJIWARA, Y., ALT, F. W. & ORKIN, S. H. 1996. The T Cell Leukemia Oncoprotein SCL/tal-1 Is Essential for Development of All Hematopoietic Lineages. *Cell,* 86**,** 47-57.

PRIVITERA, E., LONGONI, D., BRAMBILLASCA, F. & BIONDI, A. 1997. EVI-1 gene expression in myeloid clonogenic cells from juvenile myelomonocytic leukemia (JMML). *Leukemia,* 11**,** 2045-2048.

PRONK, C. J. H., ROSSI, D. J., MÅNSSON, R., ATTEMA, J. L., NORDDAHL, G. L., CHAN, C. K. F., SIGVARDSSON, M., WEISSMAN, I. L. & BRYDER, D. 2007. Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell,* 1**,** 428-442.

PTASINSKA, A., ASSI, S. A., MANNARI, D., JAMES, S. R., WILLIAMSON, D., DUNNE, J., HOOGENKAMP, M., WU, M., CARE, M., MCNEILL, H., CAUCHY, P., CULLEN, M., TOOZE, R. M., TENEN, D. G., YOUNG, B. D., COCKERILL, P. N., WESTHEAD, D. R., HEIDENREICH, O. & BONIFER, C. 2012. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia,* 26**,** 1829-1841.

PTASINSKA, A., ASSI, SALAM A., MARTINEZ-SORIA, N., IMPERATO, MARIA R., PIPER, J., CAUCHY, P., PICKIN, A., JAMES, SALLY R., HOOGENKAMP, M., WILLIAMSON, D., WU, M., TENEN, DANIEL G., OTT, S., WESTHEAD, DAVID R., COCKERILL, PETER N., HEIDENREICH, O. & BONIFER, C. 2014. Identification of a Dynamic Core Transcriptional Network in t(8;21) AML that Regulates Differentiation Block and Self-Renewal. *Cell Reports,* 8**,** 1974-1988.

PUTZ, G., ROSNER, A., NUESSLEIN, I., SCHMITZ, N. & BUCHHOLZ, F. 2006. AML1 deletion in adult mice causes splenomegaly and lymphomas. *Oncogene,* 25**,** 929-39.

QUEK, L., OTTO, G. W., GARNETT, C., LHERMITTE, L., KARAMITROS, D., STOILOVA, B., LAU, I. J., DOONDEEA, J., USUKHBAYAR, B., KENNEDY, A., METZNER, M., GOARDON, N., IVEY, A., ALLEN, C., GALE, R., DAVIES, B., STERNBERG, A., KILLICK, S., HUNTER, H., CAHALIN, P., PRICE, A., CARR, A., GRIFFITHS, M., VIRGO, P., MACKINNON, S., GRIMWADE, D., FREEMAN, S., RUSSELL, N., CRADDOCK, C., MEAD, A., PENIKET, A., PORCHER, C. & VYAS, P. 2016. Genetically distinct leukemic stem cells in human CD34−acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *The Journal of Experimental Medicine***,** jem.20151775.

RAMPAL, R., ALKALIN, A., MADZO, J., VASANTHAKUMAR, A., PRONIER, E., PATEL, J., LI, Y., AHN, J., ABDEL-WAHAB, O., SHIH, A., LU, C., WARD, P. S., TSAI, J. J., HRICIK, T., TOSELLO, V., TALLMAN, J. E., ZHAO, X., DANIELS, D., DAI, Q., CIMINIO, L., AIFANTIS, I., HE, C., FUKS, F., TALLMAN, M. S., FERRANDO, A., NIMER, S., PAIETTA, E., THOMPSON, C. B., LICHT, J. D., MASON, C. E., GODLEY, L. A., MELNICK, A., FIGUEROA, M. E. & LEVINE, R. L. 2014. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep,* 9**,** 1841-55.

RAMPAL, R. & FIGUEROA, M. E. 2016. Wilms tumor 1 mutations in the pathogenesis of acute myeloid leukemia. *Haematologica,* 101**,** 672-679.

REEBYE, V., SAETROM, P., MINTZ, P. J., HUANG, K. W., SWIDERSKI, P., PENG, L., LIU, C., LIU, X., LINDKAER-JENSEN, S., ZACHAROULIS, D., KOSTOMITSOPOULOS, N., KASAHARA, N., NICHOLLS, J. P., JIAO, L. R., PAI, M., SPALDING, D. R., MIZANDARI, M., CHIKOVANI, T., EMARA, M. M., HAOUDI, A., TOMALIA, D. A., ROSSI, J. J. & HABIB, N. A. 2014. Novel RNA oligonucleotide improves liver function and inhibits liver carcinogenesis in vivo. *Hepatology,* 59**,** 216-27.

REINKE, H. & HÖRZ, W. 2003. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Molecular Cell,* 11**,** 1599-1607.

RIEGER, M. A., HOPPE, P. S., SMEJKAL, B. M., EITELHUBER, A. C. & SCHROEDER, T. 2009. Hematopoietic cytokines can instruct lineage choice. *Science,* 325**,** 217-8.

RODRIGUES, N. P., JANZEN, V., FORKERT, R., DOMBKOWSKI, D. M., BOYD, A. S., ORKIN, S. H., ENVER, T., VYAS, P. & SCADDEN, D. T. 2005. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood,* 106**,** 477-484.

RODRIGUES, N. P., TIPPING, A. J., WANG, Z. & ENVER, T. 2012. GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *Int J Biochem Cell Biol,* 44**,** 457-60.

ROE, J. S., MERCAN, F., RIVERA, K., PAPPIN, D. J. & VAKOC, C. R. 2015. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol Cell,* 58**,** 1028-39.

ROSS, M. E., MAHFOUZ, R., ONCIU, M., LIU, H. C., ZHOU, X., SONG, G., SHURTLEFF, S. A., POUNDS, S., CHENG, C., MA, J., RIBEIRO, R. C., RUBNITZ, J. E., GIRTMAN, K., WILLIAMS, W. K., RAIMONDI, S. C., LIANG, D. C., SHIH, L. Y., PUI, C. H. & DOWNING, J. R. 2004. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood,* 104**,** 3679-87.

ROULOIS, D., LOO YAU, H., SINGHANIA, R., WANG, Y., DANESH, A., SHEN, S. Y., HAN, H., LIANG, G., JONES, P. A., PUGH, T. J., O'BRIEN, C. & DE CARVALHO, D. D. 2015. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell,* 162**,** 961-73.

RUBIN, C. M., LARSON, R. A., ANASTASI, J., WINTER, J. N., THANGAVELU, M., VARDIMAN, J. W., ROWLEY, J. D. & LE BEAU, M. M. 1990. t(3;21)(q26;q22): a recurring chromosomal abnormality in therapy-related myelodysplastic syndrome and acute myeloid leukemia. *Blood,* 76**,** 2594-2598.

RUBIN, C. M., LARSON, R. A., BITTER, M. A., CARRINO, J. J., LE BEAU, M. M., DIAZ, M. O. & ROWLEY, J. D. 1987. Association of a chromosomal 3;21 translocation with the blast phase of chronic myelogenous leukemia. *Blood,* 70**,** 1338-1342.

RUSSELL, M., LIST, A., GREENBERG, P., WOODWARD, S., GLINSMANN, B., PARGANAS, E., IHLE, J. & TAETLE, R. 1994. Expression of EVI1 in myelodysplastic syndromes and other hematologic malignancies without 3q26 translocations. *Blood,* 84**,** 1243-1248.

SAEED, A. I., BHAGABATI, N. K., BRAISTED, J. C., LIANG, W., SHAROV, V., HOWE, E. A., LI, J., THIAGARAJAN, M., WHITE, J. A. & QUACKENBUSH, J. 2006. TM4 microarray software suite. *Methods Enzymol,* 411**,** 134-93.

SAITO, Y., KITAMURA, H., HIJIKATA, A., TOMIZAWA-MURASAWA, M., TANAKA, S., TAKAGI, S., UCHIDA, N., SUZUKI, N., SONE, A., NAJIMA, Y., OZAWA, H., WAKE, A., TANIGUCHI, S., SHULTZ, L. D., OHARA, O. & ISHIKAWA, F. 2010. Identification of therapeutic targets for quiescent, chemotherapy-resistant human leukemia stem cells. *Sci Transl Med,* 2**,** 17ra9.

SCHESSL, C., RAWAT, V. P. S., CUSAN, M., DESHPANDE, A., KOHL, T. M., ROSTEN, P. M., SPIEKERMANN, K., HUMPHRIES, R. K., SCHNITTGER, S., KERN, W., HIDDEMANN, W., QUINTANILLA-MARTINEZ, L., BOHLANDER, S. K., FEURING-BUSKE, M. & BUSKE, C. 2005. The AML1-ETO fusion gene and the FLT3 length mutation

collaborate in inducing acute leukemia in mice. *The Journal of Clinical Investigation,* 115**,** 2159-2168.

SCHNABEL, C. A., JACOBS, Y. & CLEARY, M. L. 2000. HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis. *Oncogene,* 19**,** 608-16.

SCHWABISH, M. A. & STRUHL, K. 2006. Asf1 Mediates Histone Eviction and Deposition during Elongation by RNA Polymerase II. *Molecular Cell,* 22**,** 415-422.

SCHWIEGER, M., LÖHLER, J., FRIEL, J., SCHELLER, M., HORAK, I. & STOCKING, C. 2002. AML1-ETO Inhibits Maturation of Multiple Lymphohematopoietic Lineages and Induces Myeloblast Transformation in Synergy with ICSBP Deficiency. *The Journal of Experimental Medicine,* 196**,** 1227-1240.

SELTH, L. A., SIGURDSSON, S. & SVEJSTRUP, J. Q. 2010. Transcript Elongation by RNA Polymerase II. *Annu Rev Biochem,* 79**,** 271-293.

SENYUK, V., LI, D., ZAKHAROV, A., MIKHAIL, F. M. & NUCIFORA, G. 2005. The distal zinc finger domain of AML1/MDS1/EVI1 is an oligomerization domain involved in induction of hematopoietic differentiation defects in primary cells in vitro. *Cancer Res.,* 65**,** 7603-7611.

SENYUK, V., SINHA, K. K., LI, D., RINALDI, C. R., YANAMANDRA, S. & NUCIFORA, G. 2007. Repression of RUNX1 activity by EVI1: a new role of EVI1 in leukemogenesis. *Cancer Res.,* 67**,** 5658-5666.

SHAFER, D. & GRANT, S. 2016. Update on rational targeted therapy in AML. *Blood Reviews*.

SHEN, L., ZHU, J., CHEN, F., LIN, W., CAI, J., ZHONG, J. & ZHONG, H. 2015. RUNX1-Evi-1 fusion gene inhibited differentiation and apoptosis in myelopoiesis: an in vivo study. *BMC Cancer,* 15**,** 1-13.

SHI, Y., SAWADA, J.-I., SUI, G., AFFAR, E. B., WHETSTINE, J. R., LAN, F., OGAWA, H., PO-SHAN LUKE, M., NAKATANI, Y. & SHI, Y. 2003. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature,* 422**,** 735-738.

SHILATIFARD, A. 2006. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem,* 75**,** 243-69.

SHLUSH, L. I., ZANDI, S., MITCHELL, A., CHEN, W. C., BRANDWEIN, J. M., GUPTA, V., KENNEDY, J. A., SCHIMMER, A. D., SCHUH, A. C., YEE, K. W., MCLEOD, J. L., DOEDENS, M., MEDEIROS, J. J., MARKE, R., KIM, H. J., LEE, K., MCPHERSON, J. D., HUDSON, T. J., BROWN, A. M., YOUSIF, F., TRINH, Q. M., STEIN, L. D., MINDEN, M. D., WANG, J. C. & DICK, J. E. 2014. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature,* 506**,** 328-33.

SHOGREN-KNAAK, M., ISHII, H., SUN, J.-M., PAZIN, M. J., DAVIE, J. R. & PETERSON, C. L. 2006. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science,* 311**,** 844-847.

SLOVAK, M. L., KOPECKY, K. J., CASSILETH, P. A., HARRINGTON, D. H., THEIL, K. S., MOHAMED, A., PAIETTA, E., WILLMAN, C. L., HEAD, D. R., ROWE, J. M., FORMAN, S. J. & APPELBAUM, F. R. 2000. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a

Southwest Oncology Group/Eastern Cooperative Oncology Group study. *Blood,* 96**,** 4075-4083.

SOMMER, C. A., STADTFELD, M., MURPHY, G. J., HOCHEDLINGER, K., KOTTON, D. N. & MOSTOSLAVSKY, G. 2009. iPS Cell Generation Using a Single Lentiviral Stem Cell Cassette. *Stem cells (Dayton, Ohio),* 27**,** 543-549.

SONG, W. J., SULLIVAN, M. G., LEGARE, R. D., HUTCHINGS, S., TAN, X., KUFRIN, D., RATAJCZAK, J., RESENDE, I. C., HAWORTH, C., HOCK, R., LOH, M., FELIX, C., ROY, D. C., BUSQUE, L., KURNIT, D., WILLMAN, C., GEWIRTZ, A. M., SPECK, N. A., BUSHWELLER, J. H., LI, F. P., GARDINER, K., PONCZ, M., MARIS, J. M. & GILLILAND, D. G. 1999. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet,* 23**,** 166-175.

SOOD, R., TALWAR-TRIKHA, A., CHAKRABARTI, S. R. & NUCIFORA, G. 1999. MDS1/EVI1 enhances TGF-beta1 signaling and strengthens its growth-inhibitory effect but the leukemia-associated fusion protein AML1/MDS1/EVI1, product of the t(3;21), abrogates growth-inhibition in response to TGF-beta1. *Leukemia,* 13**,** 348-357.

SPANGRUDE, G. J., HEIMFELD, S. & WEISSMAN, I. L. 1988. Purification and Characterization of Mouse Hematopoietic Stem Cells. *Science,* 241**,** 58-62.

SPECK, N. A. & BALTIMORE, D. 1987. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol,* 7**,** 1101-10.

SPECK, N. A. & GILLILAND, D. G. 2002. Core-binding factors in haematopoiesis and leukaemia. *Nat Rev Cancer,* 2**,** 502-13.

SPENSBERGER, D. & DELWEL, R. 2008. A novel interaction between the proto-oncogene Evi1 and histone methyltransferases, SUV39H1 and G9a. *FEBS Letters,* 582**,** 2761-2767.

STAVROPOULOU, V., KASPAR, S., BRAULT, L., SANDERS, MATHIJS A., JUGE, S., MORETTINI, S., TZANKOV, A., IACOVINO, M., LAU, I. J., MILNE, THOMAS A., ROYO, H., KYBA, M., VALK, PETER J. M., PETERS, ANTOINE H. F. M. & SCHWALLER, J. 2016. MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. *Cancer cell,* 30**,** 43-58.

STEIN, E. M., ALTMAN, J. K., COLLINS, R., DEANGELO, D. J., FATHI, A. T., FLINN, I., FRANKEL, A., LEVINE, R. L., MEDEIROS, B. C., PATEL, M., POLLYEA, D. A., ROBOZ, G. J., STONE, R. M., SWORDS, R. T., TALLMAN, M. S., AGRESTA, S., FAN, B., YANG, H., YEN, K. & DE BOTTON, S. 2014. AG-221, an Oral, Selective, First-in-Class, Potent Inhibitor of the IDH2 Mutant Metabolic Enzyme, Induces Durable Remissions in a Phase I Study in Patients with IDH2 Mutation Positive Advanced Hematologic Malignancies. *Blood,* 124**,** 115-115.

STERNECK, E., PAYLOR, R., JACKSON-LEWIS, V., LIBBEY, M., PRZEDBORSKI, S., TESSAROLLO, L., CRAWLEY, J. N. & JOHNSON, P. F. 1998. Selectively enhanced contextual fear conditioning in mice lacking the transcriptional regulator CCAAT/enhancer binding protein δ. *Proceedings of the National Academy of Sciences,* 95**,** 10908-10913.

STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences,* 102**,** 15545-15550.

SUN, J., RAMOS, A., CHAPMAN, B., JOHNNIDIS, J. B., LE, L., HO, Y. J., KLEIN, A., HOFMANN, O. & CAMARGO, F. D. 2014. Clonal dynamics of native haematopoiesis. *Nature,* 514**,** 322-7.

SUN, W. & DOWNING, J. R. 2004. Haploinsufficiency of AML1 results in a decrease in the number of LTR-HSCs while simultaneously inducing an increase in more mature progenitors. *Blood,* 104**,** 3565-3572.

SUN, X. J., WANG, Z., WANG, L., JIANG, Y., KOST, N., SOONG, T. D., CHEN, W. Y., TANG, Z., NAKADAI, T., ELEMENTO, O., FISCHLE, W., MELNICK, A., PATEL, D. J., NIMER, S. D. & ROEDER, R. G. 2013. A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature,* 500**,** 93-7.

SUNG, M. H., BAEK, S. & HAGER, G. L. 2016. Genome-wide footprinting: ready for prime time? *Nat Methods,* 13**,** 222-8.

SWERDLOW, S. H., CAMPO, E., HARRIS, N.L., JAFFE, E.S., PILERI, S.A., STEIN, H., THIELE, J., VARDIMAN, J.W 2008. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition*, IARC.

TAGOH, H., HIMES, R., CLARKE, D., LEENEN, P. J., RIGGS, A. D., HUME, D. & BONIFER, C. 2002. Transcription factor complex formation and chromatin fine structure alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes Dev,* 16**,** 1721-37.

TAHIROV, T. H., INOUE-BUNGO, T., MORII, H., FUJIKAWA, A., SASAKI, M., KIMURA, K., SHIINA, M., SATO, K., KUMASAKA, T., YAMAMOTO, M., ISHII, S. & OGATA, K. 2001. Structural Analyses of DNA Recognition by the AML1/Runx-1 Runt Domain and Its Allosteric Control by CBF¦. *Cell,* 104**,** 755-767.

TAKESHITA, M., ICHIKAWA, M., NITTA, E., GOYAMA, S., ASAI, T., OGAWA, S., CHIBA, S. & KUROKAWA, M. 2008. AML1-Evi-1 specifically transforms hematopoietic stem cells through fusion of the entire Evi-1 sequence to AML1. *Leukemia,* 22**,** 1241-1249.

TAM, W. L. & WEINBERG, R. A. 2013. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med,* 19**,** 1438-1449.

TANAKA, K., TANAKA, T., KUROKAWA, M., IMAI, Y., OGAWA, S., MITANI, K., YAZAKI, Y. & HIRAI, H. 1998. The AML1/ETO(MTG8) and AML1/Evi-1 leukemia-associated chimeric oncoproteins accumulate PEBP2beta(CBFbeta) in the nucleus more efficiently than wild-type AML1. *Blood,* 91**,** 1688-1699.

TANAKA, T., MITANI, K., KUROKAWA, M., OGAWA, S., TANAKA, K., NISHIDA, J., YAZAKI, Y., SHIBATA, Y. & HIRAI, H. 1995. Dual functions of the AML1/Evi-1 chimeric protein in the mechanism of leukemogenesis in t(3;21) leukemias. *Mol.Cell Biol.,* 15**,** 2383-2392.

TANAKA, T., NISHIDA, J., MITANI, K., OGAWA, S., YAZAKI, Y. & HIRAI, H. 1994. Evi-1 raises AP-1 activity and stimulates c-fos promoter transactivation with dependence on the second zinc finger domain. *J.Biol.Chem.,* 269**,** 24020-24026.

TANAKA, T., YOSHIDA, N., KISHIMOTO, T. & AKIRA, S. 1997. Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. *EMBO J,* 16**,** 7432-43.

TANIUCHI, I., OSATO, M., EGAWA, T., SUNSHINE, M. J., BAE, S. C., KOMORI, T., ITO, Y. & LITTMAN, D. R. 2002. Differential Requirements for Runx Proteins in CD4 Repression and Epigenetic Silencing during T Lymphocyte Development. *Cell,* 111**,** 621-633.

TAUSSIG, D. C., VARGAFTIG, J., MIRAKI-MOUD, F., GRIESSINGER, E., SHARROCK, K., LUKE, T., LILLINGTON, D., OAKERVEE, H., CAVENAGH, J., AGRAWAL, S. G., LISTER, T. A., GRIBBEN, J. G. & BONNET, D. 2010. Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34− fraction. *Blood,* 115**,** 1976-1984.

TAVOR, S., VUONG, P. T., PARK, D. J., GOMBART, A. F., COHEN, A. H. & KOEFFLER, H. P. 2002. Macrophage functional maturation and cytokine production are impaired in C/EBPε-deficient mice. *Blood,* 99**,** 1794-1801.

THORNELL, A., HALLBERG, B. & GRUNDSTROM, T. 1988. Differential protein binding in lymphocytes to a sequence in the enhancer of the mouse retrovirus SL3-3. *Mol Cell Biol,* 8**,** 1625-37.

THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., EBERSOL, A. K., FRUM, T., GISTE, E., JOHNSON, A. K., JOHNSON, E. M., KUTYAVIN, T., LAJOIE, B., LEE, B.-K., LEE, K., LONDON, D., LOTAKIS, D., NEPH, S., NERI, F., NGUYEN, E. D., QU, H., REYNOLDS, A. P., ROACH, V., SAFI, A., SANCHEZ, M. E., SANYAL, A., SHAFER, A., SIMON, J. M., SONG, L., VONG, S., WEAVER, M., YAN, Y., ZHANG, Z., ZHANG, Z., LENHARD, B., TEWARI, M., DORSCHNER, M. O., HANSEN, R. S., NAVAS, P. A., STAMATOYANNOPOULOS, G., IYER, V. R., LIEB, J. D., SUNYAEV, S. R., AKEY, J. M., SABO, P. J., KAUL, R., FUREY, T. S., DEKKER, J., CRAWFORD, G. E. & STAMATOYANNOPOULOS, J. A. 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

TIJSSEN, M. R., CVEJIC, A., JOSHI, A., HANNAH, R. L., FERREIRA, R., FORRAI, A., BELLISSIMO, D. C., ORAM, S. H., SMETHURST, P. A., WILSON, N. K., WANG, X., OTTERSBACH, K., STEMPLE, D. L., GREEN, A. R., OUWEHAND, W. H. & GOTTGENS, B. 2011. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell,* 20**,** 597-609.

TIMCHENKO, L. T., IAKOVA, P., WELM, A. L., CAI, Z.-J. & TIMCHENKO, N. A. 2002. Calreticulin Interacts with C/EBPα and C/EBPβ mRNAs and Represses Translation of C/EBP Proteins. *Molecular and Cellular Biology,* 22**,** 7242-7257.

TIMCHENKO, N., WILSON, D. R., TAYLOR, L. R., ABDELSAYED, S., WILDE, M., SAWADOGO, M. & DARLINGTON, G. J. 1995. Autoregulation of the human

C/EBP alpha gene by stimulation of upstream stimulatory factor binding. *Mol Cell Biol,* 15**,** 1192-202.

TIPPING, A. J., PINA, C., CASTOR, A., HONG, D., RODRIGUES, N. P., LAZZARI, L., MAY, G. E., JACOBSEN, S. E. W. & ENVER, T. 2009. High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood,* 113**,** 2661-2672.

TOKITA, K., MAKI, K. & MITANI, K. 2007. RUNX1/EVI1, which blocks myeloid differentiation, inhibits CCAAT-enhancer binding protein alpha function. *Cancer Sci.,* 98**,** 1752-1757.

TSAI, F.-Y., KELLER, G., KUO, F. C., WEISS, M., CHEN, J., ROSENBLATT, M., ALT, F. W. & ORKIN, S. H. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature,* 371**,** 221-226.

TURNER, B. M. 2002. Cellular Memory and the Histone Code. *Cell,* 111**,** 285-291.

TURNER, B. M. 2012. The adjustable nucleosome: an epigenetic signaling module. *Trends in Genetics,* 28**,** 436-444.

TURSKY, M. L., BECK, D., THOMS, J. A., HUANG, Y., KUMARI, A., UNNIKRISHNAN, A., KNEZEVIC, K., EVANS, K., RICHARDS, L. A., LEE, E., MORRIS, J., GOLDBERG, L., IZRAELI, S., WONG, J. W., OLIVIER, J., LOCK, R. B., MACKENZIE, K. L. & PIMANDA, J. E. 2015. Overexpression of ERG in cord blood progenitors promotes expansion and recapitulates molecular signatures of high ERG leukemias. *Leukemia,* 29**,** 819-27.

UMEK, R. M., FRIEDMAN, A. D. & MCKNIGHT, S. L. 1991. CCAAT-enhancer binding protein: A component of a differentiation switch. *Science,* 251**,** 288-292.

VALK, P. J., VERHAAK, R. G., BEIJEN, M. A., ERPELINCK, C. A., BARJESTEH VAN WAALWIJK VAN DOORN-KHOSROVANI, S., BOER, J. M., BEVERLOO, H. B., MOORHOUSE, M. J., VAN DER SPEK, P. J., LOWENBERG, B. & DELWEL, R. 2004. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med,* 350**,** 1617-28.

VAN OEVELEN, C., COLLOMBET, S., VICENT, G., HOOGENKAMP, M., LEPOIVRE, C., BADEAUX, A., BUSSMANN, L., SARDINA, J. L., THIEFFRY, D., BEATO, M., SHI, Y., BONIFER, C. & GRAF, T. 2015. C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports,* 5**,** 232-47.

VERHAAK, R. G. W., WOUTERS, B. J., ERPELINCK, C. A. J., ABBAS, S., BEVERLOO, H. B., LUGTHART, S., LÖWENBERG, B., DELWEL, R. & VALK, P. J. M. 2009. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica,* 94**,** 131-134.

VIERSTRA, J. & STAMATOYANNOPOULOS, J. A. 2016. Genomic footprinting. *Nat Methods,* 13**,** 213-21.

VINATZER, U., TAPLICK, J., SEISER, C., FONATSCH, C. & WIESER, R. 2001. The leukaemia-associated transcription factors EVI-1 and MDS1/EVI1 repress transcription and interact with histone deacetylase. *Br.J.Haematol.,* 114**,** 566-573.

WANG, J., HOSHINO, T., REDNER, R. L., KAJIGAYA, S. & LIU, J. M. 1998. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by

interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc Natl Acad Sci U S A,* 95**,** 10860-5.

WANG, Q., STACY, T., MILLER, J. D., LEWIS, A. F., GU, T. L., HUANG, X., BUSHWELLER, J. H., BORIES, J. C., ALT, F. W., RYAN, G., LIU, P. P., WYNSHAW-BORIS, A., BINDER, M., MAR+¡N-PADILLA, M., SHARPE, A. H. & SPECK, N. A. 1996. The CBF+¦ Subunit Is Essential for CBF+¦2 (AML1) Function In Vivo. *Cell,* 87**,** 697-708.

WANG, Y., XIAO, M., CHEN, X., CHEN, L., XU, Y., LV, L., WANG, P., YANG, H., MA, S., LIN, H., JIAO, B., REN, R., YE, D., GUAN, K. L. & XIONG, Y. 2015. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Mol Cell,* 57**,** 662-73.

WARNIER, M., ROUDBARAKI, M., DEROUICHE, S., DELCOURT, P., BOKHOBZA, A., PREVARSKAYA, N. & MARIOT, P. 2015. CACNA2D2 promotes tumorigenesis by stimulating cell proliferation and angiogenesis. *Oncogene,* 34**,** 5383-94.

WEINTRAUB, H. & GROUDINE, M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science,* 193**,** 848-856.

WEISSMANN, S., ALPERMANN, T., GROSSMANN, V., KOWARSCH, A., NADARAJAH, N., EDER, C., DICKER, F., FASAN, A., HAFERLACH, C., HAFERLACH, T., KERN, W., SCHNITTGER, S. & KOHLMANN, A. 2012. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia,* 26**,** 934-942.

WELCH, JOHN S., LEY, TIMOTHY J., LINK, DANIEL C., MILLER, CHRISTOPHER A., LARSON, DAVID E., KOBOLDT, DANIEL C., WARTMAN, LUKAS D., LAMPRECHT, TAMARA L., LIU, F., XIA, J., KANDOTH, C., FULTON, ROBERT S., MCLELLAN, MICHAEL D., DOOLING, DAVID J., WALLIS, JOHN W., CHEN, K., HARRIS, CHRISTOPHER C., SCHMIDT, HEATHER K., KALICKI-VEIZER, JOELLE M., LU, C., ZHANG, Q., LIN, L., O'LAUGHLIN, MICHELLE D., MCMICHAEL, JOSHUA F., DELEHAUNTY, KIM D., FULTON, LUCINDA A., MAGRINI, VINCENT J., MCGRATH, SEAN D., DEMETER, RYAN T., VICKERY, TAMMI L., HUNDAL, J., COOK, LISA L., SWIFT, GARY W., REED, JERRY P., ALLDREDGE, PATRICIA A., WYLIE, TODD N., WALKER, JASON R., WATSON, MARK A., HEATH, SHARON E., SHANNON, WILLIAM D., VARGHESE, N., NAGARAJAN, R., PAYTON, JACQUELINE E., BATY, JACK D., KULKARNI, S., KLCO, JEFFERY M., TOMASSON, MICHAEL H., WESTERVELT, P., WALTER, MATTHEW J., GRAUBERT, TIMOTHY A., DIPERSIO, JOHN F., DING, L., MARDIS, ELAINE R. & WILSON, RICHARD K. 2012. The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell,* 150**,** 264-278.

WESTENDORF, J. J., YAMAMOTO, C. M., LENNY, N., DOWNING, J. R., SELSTED, M. E. & HIEBERT, S. W. 1998. The t(8;21) Fusion Product, AML-1ΓÇôETO, Associates with C/EBP-+¦, Inhibits C/EBP-+¦-Dependent Transcription, and Blocks Granulocytic Differentiation. *Molecular and Cellular Biology,* 18**,** 322-333.

WICHMANN, C., QUAGLIANO-LO COCO, I., YILDIZ, O., CHEN-WICHMANN, L., WEBER, H., SYZONENKO, T., DORING, C., BRENDEL, C., PONNUSAMY, K., KINNER, A., BRANDTS, C., HENSCHLER, R. & GREZ, M. 2015. Activating c-KIT mutations confer oncogenic cooperativity and rescue RUNX1/ETO-induced DNA damage and apoptosis in human primary CD34+ hematopoietic progenitors. *Leukemia,* 29**,** 279-89.

WIEMELS, J. L., XIAO, Z., BUFFLER, P. A., MAIA, A. T., MA, X., DICKS, B. M., SMITH, M. T., ZHANG, L., FEUSNER, J., WIENCKE, J., PRITCHARD-JONES, K., KEMPSKI, H. & GREAVES, M. 2002. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood,* 99**,** 3801-5.

WILSON, A., LAURENTI, E., OSER, G., VAN DER WATH, R. C., BLANCO-BOSE, W., JAWORSKI, M., OFFNER, S., DUNANT, C. F., ESHKIND, L., BOCKAMP, E., LIO, P., MACDONALD, H. R. & TRUMPP, A. 2008. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell,* 135**,** 1118-29.

WILSON, N. K., FOSTER, S. D., WANG, X., KNEZEVIC, K., SCHUTTE, J., KAIMAKIS, P., CHILARSKA, P. M., KINSTON, S., OUWEHAND, W. H., DZIERZAK, E., PIMANDA, J. E., DE BRUIJN, M. F. & GOTTGENS, B. 2010. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell,* 7**,** 532-44.

WONG, P., IWASAKI, M., SOMERVAILLE, T. C. P., SO, C. W. E. & CLEARY, M. L. 2007. Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. *Genes & Development,* 21**,** 2762-2774.

WONG, T. N., RAMSINGH, G., YOUNG, A. L., MILLER, C. A., TOUMA, W., WELCH, J. S., LAMPRECHT, T. L., SHEN, D., HUNDAL, J., FULTON, R. S., HEATH, S., BATY, J. D., KLCO, J. M., DING, L., MARDIS, E. R., WESTERVELT, P., DIPERSIO, J. F., WALTER, M. J., GRAUBERT, T. A., LEY, T. J., DRULEY, T. E., LINK, D. C. & WILSON, R. K. 2015. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature,* 518**,** 552-555.

WOTTON, D., GHYSDAEL, J., WANG, S., SPECK, N. A. & OWEN, M. J. 1994. Cooperative binding of Ets-1 and core binding factor to DNA. *Mol Cell Biol,* 14**,** 840-50.

WU, C. 1980. The 5[prime] ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature,* 286**,** 854-860.

WU, H. & ZHANG, Y. 2014. Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell,* 156**,** 45-68.

XIANG, P., WEI, W., LO, C., ROSTEN, P., HOU, J., HOODLESS, P. A., BILENKY, M., BONIFER, C., COCKERILL, P. N., KIRKPATRICK, A., GOTTGENS, B., HIRST, M. & HUMPHRIES, K. R. 2014. Delineating MEIS1 cis-regulatory elements active in hematopoietic cells. *Leukemia,* 28**,** 433-6.

XIE, H., YE, M., FENG, R. & GRAF, T. 2004. Stepwise Reprogramming of B Cells into Macrophages. *Cell,* 117**,** 663-676.

XU, F., WU, L.-Y., CHANG, C.-K., HE, Q., ZHANG, Z., LIU, L., SHI, W.-H., GUO, J., ZHU, Y., ZHAO, Y.-S., GU, S.-C., FEI, C.-M., WU, D., ZHOU, L.-Y., SU, J.-Y., SONG, L.-X., XIAO, C. & LI, X. 2015. Whole-exome and targeted sequencing identify ROBO1 and ROBO2 mutations as progression-related drivers in myelodysplastic syndromes. *Nat Commun,* 6.

YAGI, T., MORIMOTO, A., EGUCHI, M., HIBI, S., SAKO, M., ISHII, E., MIZUTANI, S., IMASHUKU, S., OHKI, M. & ICHIKAWA, H. 2003. Identification of a gene expression signature associated with pediatric AML prognosis. *Blood,* 102**,** 1849-56.

YAMANAKA, R., BARLOW, C., LEKSTROM-HIMES, J., CASTILLA, L. H., LIU, P. P., ECKHAUS, M., DECKER, T., WYNSHAW-BORIS, A. & XANTHOPOULOS, K. G. 1997. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein ε-deficient mice. *Proceedings of the National Academy of Sciences,* 94**,** 13187-13192.

YAN, P., FRANKHOUSER, D., MURPHY, M., TAM, H. H., RODRIGUEZ, B., CURFMAN, J., TRIMARCHI, M., GEYER, S., WU, Y. Z., WHITMAN, S. P., METZELER, K., WALKER, A., KLISOVIC, R., JACOB, S., GREVER, M. R., BYRD, J. C., BLOOMFIELD, C. D., GARZON, R., BLUM, W., CALIGIURI, M. A., BUNDSCHUH, R. & MARCUCCI, G. 2012. Genome-wide methylation profiling in decitabine-treated patients with acute myeloid leukemia. *Blood,* 120**,** 2466-74.

YANG, L., RODRIGUEZ, B., MAYLE, A., PARK, HYUN J., LIN, X., LUO, M., JEONG, M., CURRY, CHOLADDA V., KIM, S.-B., RUAU, D., ZHANG, X., ZHOU, T., ZHOU, M., REBEL, VIVIENNE I., CHALLEN, GRANT A., GÖTTGENS, B., LEE, J.-S., RAU, R., LI, W. & GOODELL, MARGARET A. 2016. DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. *Cancer cell,* 29**,** 922-934.

YANG, X. J., OGRYZKO, V. V., NISHIKAWA, J., HOWARD, B. H. & NAKATANI, Y. 1996. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature,* 382**,** 319-24.

YATSULA, B., LIN, S., READ, A. J., POHOLEK, A., YATES, K., YUE, D., HUI, P. & PERKINS, A. S. 2005. Identification of binding sites of EVI1 in mammalian cells. *J Biol Chem,* 280**,** 30712-22.

YE, M., ZHANG, H., AMABILE, G., YANG, H., STABER, P. B., ZHANG, P., LEVANTINI, E., ALBERICH-JORDA, M., ZHANG, J., KAWASAKI, A. & TENEN, D. G. 2013. C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. *Nat Cell Biol,* 15**,** 385-94.

YERGEAU, D. A., HETHERINGTON, C. J., WANG, Q., ZHANG, P., SHARPE, A. H., BINDER, M., MARIN-PADILLA, M., TENEN, D. G., SPECK, N. A. & ZHANG, D. E. 1997. Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene. *Nat Genet,* 15**,** 303-306.

YOSHIMI, A., GOYAMA, S., WATANABE-OKOCHI, N., YOSHIKI, Y., NANNYA, Y., NITTA, E., ARAI, S., SATO, T., SHIMABE, M., NAKAGAWA, M., IMAI, Y., KITAMURA, T. & KUROKAWA, M. 2011. Evi1 represses PTEN expression and activates PI3K/AKT/mTOR via interactions with polycomb proteins. *Blood,* 117**,** 3617-3628.

YUAN, Y., ZHOU, L., MIYAMOTO, T., IWASAKI, H., HARAKAWA, N., HETHERINGTON, C. J., BUREL, S. A., LAGASSE, E., WEISSMAN, I. L., AKASHI, K. & ZHANG, D. E. 2001. AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc.Natl.Acad.Sci.U.S A,* 98**,** 10398-10403.

YUASA, H., OIKE, Y., IWAMA, A., NISHIKATA, I., SUGIYAMA, D., PERKINS, A., MUCENSKI, M. L., SUDA, T. & MORISHITA, K. 2005. Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression. *EMBO J.,* 24**,** 1976-1987.

ZARET, K. S. & CARROLL, J. S. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development,* 25**,** 2227-2241.

ZARET, K. S., LERNER, J. & IWAFUCHI-DOI, M. 2016. Chromatin Scanning by Dynamic Binding of Pioneer Factors. *Mol Cell,* 62**,** 665-7.

ZHANG, D.-E., ZHANG, P., WANG, N.-D., HETHERINGTON, C. J., DARLINGTON, G. J. & TENEN, D. G. 1997a. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice. *Proceedings of the National Academy of Sciences,* 94**,** 569-574.

ZHANG, D. E., FUJIOKA, K., HETHERINGTON, C. J., SHAPIRO, L. H., CHEN, H. M., LOOK, A. T. & TENEN, D. G. 1994. Identification of a region which directs the monocytic activity of the colony-stimulating factor 1 (macrophage colony-stimulating factor) receptor promoter and binds PEBP2/CBF (AML1). *Molecular and Cellular Biology,* 14**,** 8085-8095.

ZHANG, D. E., HETHERINGTON, C. J., MEYERS, S., RHOADES, K. L., LARSON, C. J., CHEN, H. M., HIEBERT, S. W. & TENEN, D. G. 1996. CCAAT enhancer-binding protein (C/EBP) and AML1 (CBF alpha2) synergistically activate the macrophage colony-stimulating factor receptor promoter. *Molecular and Cellular Biology,* 16**,** 1231-1240.

ZHANG, D. E., ZHANG, P., WANG, N. D., HETHERINGTON, C. J., DARLINGTON, G. J. & TENEN, D. G. 1997b. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci U S A,* 94**,** 569-74.

ZHANG, H., ALBERICH-JORDA, M., AMABILE, G., YANG, H., STABER, P. B., DI RUSCIO, A., WELNER, R. S., EBRALIDZE, A., ZHANG, J., LEVANTINI, E., LEFEBVRE, V., VALK, P. J., DELWEL, R., HOOGENKAMP, M., NERLOV, C., CAMMENGA, J., SAEZ, B., SCADDEN, D. T., BONIFER, C., YE, M. & TENEN, D. G. 2013. Sox4 is a key oncogenic target in C/EBPalpha mutant acute myeloid leukemia. *Cancer cell,* 24**,** 575-88.

ZHANG, J., KALKUM, M., YAMAMURA, S., CHAIT, B. T. & ROEDER, R. G. 2004a. E Protein Silencing by the Leukemogenic AML1-ETO Fusion Protein. *Science,* 305**,** 1286-1289.

ZHANG, P., IWASAKI-ARAI, J., IWASAKI, H., FENYUS, M. L., DAYARAM, T., OWENS, B. M., SHIGEMATSU, H., LEVANTINI, E., HUETTNER, C. S., LEKSTROM-HIMES, J. A., AKASHI, K. & TENEN, D. G. 2004b. Enhancement of Hematopoietic Stem Cell Repopulating Capacity and Self-Renewal in the Absence of the Transcription Factor C/EBPα. *Immunity,* 21**,** 853-863.

ZHANG, Y., STRISSEL, P., STRICK, R., CHEN, J., NUCIFORA, G., LE BEAU, M. M., LARSON, R. A. & ROWLEY, J. D. 2002. Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia. *Proc Natl Acad Sci U S A,* 99**,** 3070-5.

ZHU, H., WANG, G. & QIAN, J. 2016. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet,* 17**,** 551-65.

ZOR, T., MAYR, B. M., DYSON, H. J., MONTMINY, M. R. & WRIGHT, P. E. 2002. Roles of Phosphorylation and Helix Propensity in the Binding of the KIX Domain of CREB-

binding Protein by Constitutive (c-Myb) and Inducible (CREB) Activators. *Journal of Biological Chemistry,* 277**,** 42241-42248.

ZOVEIN, A. C., HOFMANN, J. J., LYNCH, M., FRENCH, W. J., TURLO, K. A., YANG, Y., BECKER, M. S., ZANETTA, L., DEJANA, E., GASSON, J. C., TALLQUIST, M. D. & IRUELA-ARISPE, M. L. 2008. Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells. *Cell Stem Cell,* 3**,** 625-636.

ZUBER, J., SHI, J., WANG, E., RAPPAPORT, A. R., HERRMANN, H., SISON, E. A., MAGOON, D., QI, J., BLATT, K., WUNDERLICH, M., TAYLOR, M. J., JOHNS, C., CHICAS, A., MULLOY, J. C., KOGAN, S. C., BROWN, P., VALENT, P., BRADNER, J. E., LOWE, S. W. & VAKOC, C. R. 2011. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature,* 478**,** 524-8.